# DYNAMIC COMBINATORIAL CHEMISTRY AS A TOOL IN THE IDENTIFICATION OF NOVEL RECEPTORS FOR BIOMOLECULES

Lindsey Ann Ingerman

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry.

Chapel Hill
2010

Approved by:

Professor Marcey Waters

Professor Michel Gagné

Professor Maurice Brookhart

Professor David Lawrence

Professor James Jorgenson

**Abstract**

Lindsey Ann Ingerman: Dynamic Combinatorial Chemistry as a Tool in the Identification of Novel Receptors for Biomolecules

(Under the direction of Marcey L. Waters)

The work presented in this thesis highlights various advances in the field of dynamic combinatorial chemistry (DCC), both in the development of new types of dynamic libraries and in the investigation of molecular receptors for biomolecules of interest. DCC has emerged in recent years as a new strategy for the discovery of host-guest systems based on the generation of libraries via reversible chemistry. The true utility of DCC lies in the fact that recognition of a guest molecule causes the equilibrium to shift, allowing for amplification and detection of novel receptors.

This technology has been utilized most notably for the identification of synthetic receptors for protein post-translational modifications, particularly methylated lysines and arginines. These modifications are of great interest due to their crucial role in gene expression and cell signaling. Small molecule receptors have been demonstrated via DCC that discriminate for trimethyllysine over the lower methylation states, for example, paralleling the affinity and selectivity of the native protein receptor. This suggests that such synthetic receptors have a promising future as affinity reagents for PTMs. Also of interest are methylated nucleotides, and larger peptide based macrocycles have been shown to function as selective receptors for 7-methyl guanosine via DCC, revealing the importance of the methyl group in strengthening this host-guest interaction.

In addition to executing molecular recognition studies via DCC, we have expanded upon the present collection of building blocks and libraries that have been demonstrated previously. An azobenzene moiety was incorporated into a doubly dynamic library to highlight the utility of photocontrolled libraries. Through templation with a polyproline peptide the amplification of an azobenzene containing macrocycle was demonstrated. Photoresponsive receptors of this type have the potential to allow for greater control over molecular recognition events. Furthermore, new libraries of peptidic marcocycles were demonstrated via thioester exchange. This work is particularly advantageous in that it is feasible to generate large numbers of cyclic peptides efficiently within hours for further screening. The reaction dynamics and the kinetic determinants of macrocycle formation were investigated and found to be highly dependant on the peptide building block structures.

**Acknowledgements**

First and foremost I would like to thank my advisor, Professor Marcey Waters, for her tireless support, encouragement, inspiration, and friendship. I admire her enthusiasm for her work and am extremely thankful for her mentorship over the past five years. She has undoubtedly shaped the type of scientist that I have become and for that I will always be grateful.

I would also like to thank my entire committee, Professor Michel Gagné, Professor Maurice Brookhart, Professor David Lawrence, and Professor James Jorgenson, for their guidance and assistance along the way, both in and out of the classroom. My learning experience was enhanced considerably by my interactions with my committee, and I owe them all a great deal of gratitude. I owe a special thank you to Dr. Gagné for his collaborative efforts and continual advice.

There are various other members of the UNC scientific community that I could not have succeeded without, including Dr. Ashutosh Tripathy, Dr. Marc ter Horst, Dr. David Harris, Dr. Sohrab Habibi, and Dr. Matthew Crowe. I am extremely thankful for their scientific expertise, as well as the time they each spent teaching and assisting me with a range of scientific problems. I also thank our collaborators, Professor Brian Strahl and Dr. Stephen Fuchs from the Department of Biochemistry and Biophysics.

I would like to thank the entire Waters Lab, past and present. They made it enjoyable to come to lab everyday, providing a great deal of advice and companionship, and most

importantly laughter. I can't imagine having spent these five years working with a better group of people.

I am incredibly grateful for the support I have received from my friends and family throughout this journey and I cannot thank them enough. My parents are the main reason that this thesis was possible, for they always made me dream bigger and without them I would not be the person that I am today. Last, I would like to thank my fiancé, Drew, for having more confidence in me than I will ever have in myself, and for making the past few years more exciting than I ever could have imagined.

## Table of Contents

## List of Tables

## List of Figures

## List of Schemes

## Abbreviations

| | |
|---|---|
| ACN | Acetonitrile |
| AF | Amplification factor |
| Arg, R | Arginine |
| ArgMe, RMe | Monomethylarginine |
| $ArgMe_2$, $RMe_2$ | Dimethylarginine |
| aRMe2 | Asymmetric dimethylarginine |
| Asn, N | Asparagine |
| Asp, D | Aspartic acid |
| ATP | Adenosine 5'-triphosphate |
| CD | Circular dichroism |
| $CHCl_3$ | Chloroform |
| Cys, C | Cysteine |
| DCC | Dynamic combinatorial chemistry |
| DCL | Dynamic combinatorial library |
| DCM | Dichloromethane |
| DIPEA | Diisopropylethyl amine |
| DMF | Dimethylformamide |
| DMSO | Dimethyl sulfoxide |
| DNA | Deoxyribonucleic acid |
| DSS | 3-(Trimethylsilyl)-1-propanesulfonic acid sodium salt |
| EDT | 1,2-Ethanedithiol |
| eIF4E | Eukaryotic initiation factor 4E |
| ESI | Electrospray ionization |
| FAM | Carboxyfluorescein |
| Fmoc | N-9-Fluorenylmethyloxycarbonyl |
| Gln, Q | Glutamine |
| Glu, E | Glutamic acid |
| Gly, G | Glycine |
| GMP | Guanosine 5'-monophosphate |

| | |
|---|---|
| GTP | Guanosine 5’-triphosphate |
| H3 | Histone 3 |
| HBTU | 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate |
| His, H | Histidine |
| HOBT | N-hydroxybenzotriazole |
| HP1 | Heterochromatin protein 1 |
| HPLC | High pressure liquid chromatography |
| Hyp | Hydroxyproline |
| IsoAsp | Isoaspartic acid |
| ITC | Isothermal titration calorimetry |
| LC-MS | Liquid chromatography – mass spectrometry |
| Lys, K | Lysine |
| LysMe, KMe | Monomethyllysine |
| $LysMe_2$, $KMe_2$ | Dimethyllysine |
| $LysMe_3$, $KMe_3$ | Trimethyllysine |
| MS | Mass spectrometry |
| NMR | Nuclear magnetic resonance |
| Pbf | 2,2,4,6,7-Pentamethyl-dihydrobenzofurance-5-sulfonyl |
| PEG | Polyethylene glycol |
| Phe, F | Phenylalanine |
| Pro, P | Proline |
| PTM | Post-translational modification |
| sRMe2 | Symmetric dimethylarginine |
| TCEP | Tris(2-carboxyethyl)phosphine |
| TFA | Trifluoroacetic acid |
| TFE | Trifluoroethanol |
| TIPS | Triisopropyl silane |
| Trp, W | Tryptophan |
| Trt | Trityl |
| Val, V | Valine |

# CHAPTER I

# DYNAMIC COMBINATORIAL CHEMISTRY

## A. Introduction

In attempt to mimic the thriving process of natural selection where those deemed most "fit" are rewarded with the ability to successfully reproduce, since the mid-1990s chemists have been developing their own natural selection process involving both selection and amplification, known as dynamic combinatorial chemistry (DCC). Dynamic combinatorial chemistry is an attractive alternative to traditional rational design in the development of novel receptors, in which molecular recognition guides the synthesis of complex host systems from simple building blocks using reversible linkages under thermodynamic control (Figure 1.1).[1] It takes advantage of the ability of a molecular target to template the preferential bond formation of the strongest target binders, upon which desired receptors are amplified at the expense of other oligomers.[2] DCC not only allows for the discovery of non-intuitive receptors, but it also provides access to structures that may otherwise be unobtainable by traditional synthesis.[3]

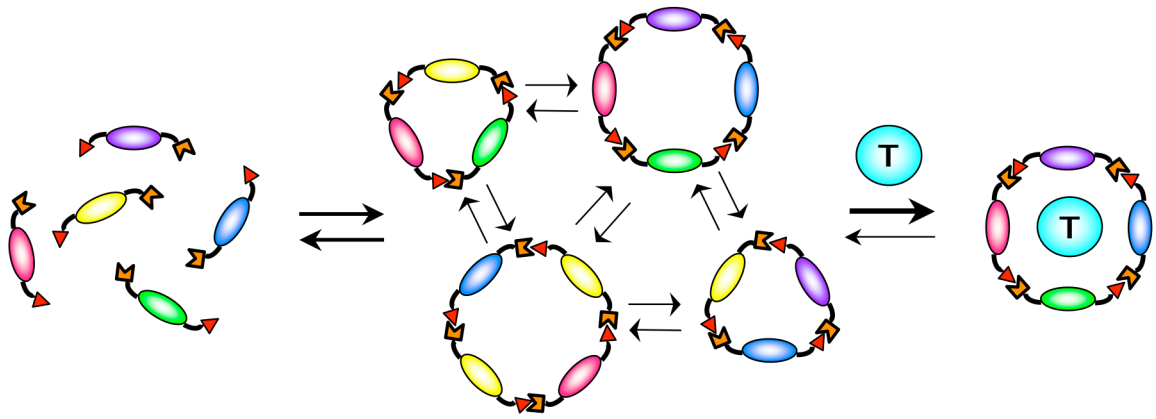



**Figure 1.1.** Generation and templation of a dynamic combinatorial library (DCL).

There are several key components of DCC that together allow for a system that can facilitate Darwinian like evolution, each of which will be introduced to some degree to provide adequate background for the work in this thesis. First, a set of designed monomers which serve as building blocks in the generation of a larger library of constituents are needed. Second, the implementation of reversible chemistry between monomers is required to generate a group of library members which are under thermodynamic control. This highlights the primary difference between traditional combinatorial chemistry and DCC, as combinatorial chemistry functions under kinetic control. Third, a selection mechanism is necessary, which is most commonly a form of molecular recognition or receptor binding, however many other selection methods have been demonstrated. Last, an analytical method is required to identify those compounds which are most "fit" and are thus are enhanced via this selection process.

**B. Reversible exchange reactions**

While the principal requirement of a reaction used to facilitate exchange between building blocks in a dynamic combinatorial library is that it be reversible, there are multiple other criteria that must be met in order for the system to serve as an effective selection process.[1a] To begin, the rate of reversible exchange is important in that it should occur on a reasonable time scale. While this requirement is often a matter of convenience, it can also be important when degradation of the target molecule over time is an issue. Next, due to the simultaneous nature of the equilibration and selection processes, the reversible reaction must proceed under conditions that are compatible with the target of interest, including all functional groups on the molecule. This becomes particularly relevant, and often challenging, in the screening of biomolecules which generally require physiological conditions (neutral pH in buffered aqueous solution). While the reaction conditions must be compatible with the experimental conditions for selection, they must also be fairly mild (for example, ambient temperature and pressure) as to not interfere with the non-covalent interactions involved in molecular recognition. The reversible reaction must also ensure the solubility of all library members formed at equilibrium, as any insoluble species can potentially act as a thermodynamic trap. Lastly, the ability to turn off the exchange reaction and kinetically "freeze" the library distribution is essential for analysis of the library distribution and to fully assess the degree of amplification. The isolation and further investigation of identified receptors is also only feasible if exchange has been halted.

Both a wide range of covalent and noncovalent reversible reactions have been demonstrated in the context of dynamic combinatorial libraries. While reversible

covalent bonds often exhibit slower kinetics, they are more widespread due to the fact that enhanced compounds can be easily isolated and handled further. In contrast, libraries formed based on noncovalent interactions achieve equilibration rapidly, yet the often weak and labile bonds generate products that are difficult to analyze in solution and even more problematic to isolate. Despite this, both hydrogen bonds[4] and metal-ligand coordination bonds[5] have been successfully applied to DCC experiments in numerous cases. While the full scope of potential reversible reactions is exhaustive and beyond the scope of this thesis, it is worth addressing those that have received significant attention thus far in this field.

Nearly all reversible unsymmetrical covalent bonds used in DCC involve carbonyl compounds or some derivative thereof (Figure 1.2). Acyl exchange reactions such as trans-esterification, allyl ester exchange, amide exchange, aldol exchange, and thioester exchange are all facilitated by the relatively fragile nature of the C(O)-X bonds, which can consequently be reversibly broken. One of the earliest examples where reversible transesterification was utilized in DCC studies was reported by Sanders and coworkers, where a cholic acid methyl ester derivative was used to form an equilibrating mixture of linear and cyclic oligomers in the presence of a potassium methoxide-crown ether complex.[6] Amide bond exchange, while slightly more challenging due to the stability of amide bonds, has also been demonstrated based on approaches which employ biocatalysts, such as proteases.[7] While important steps have been made towards the use of synthetic catalysts for transamidation reactions, the conditions employed are not yet compatible with biomolecular recognition.[8] Thioester exchange has received attention more recently as it proceeds both at room temperature and under neutral conditions, and

has been employed in the context of libraries targeting hydrolases[9] and in the evaluation of the higher order structural stability in polypeptides.[10]

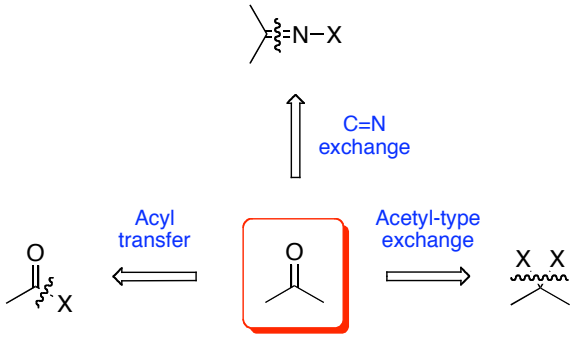



**Figure 1.2.** Different types of unsymmetrical covalent exchange reactions involving carbonyl groups: acyl transfer (X = NHR, OR, SR, $CH_2R$), C=N exchange (X = R, NHR, OR), and acetal-type exchange (X = OR, SR, NHR).

Several studies have also explored exchange reactions that benefit from the labile C=N double bond, such as imine, oxime, and hydrazone exchange. The first demonstration of imine exchange in DCC was described by Huc and Lehn in the generation of a library targeting the production of carbonic anhydrase inhibitors,[11] and imine exchange has since remained prevalent in the DCC literature. Structurally related hydrazones and oximes are often viewed as advantageous due to their great hydrolytic stability in comparison to imines. While the application of reversible oxime chemistry in the generation of libraries has been somewhat scarce since its first demonstration by Eliseev and coworkers in the late 1990s,[12] hydrazone exchange has proven to be a quite successful and powerful technique in the generation of diverse DCLs. Numerous examples have been presented based on pseudo-mono-and dipeptides in the generation of acyl hydrazone libraries,[13] many of which have resulted in receptors for cationic templates such as quaternary ammoniums[14] and alkali cations,[15] as well as nucleotide bases.[16]

Transacetalization chemistry based on the well-known formation of cyclic acetals from diols and aldehydes has been demonstrated in relatively few studies in comparison to acyl and C=N exchange reactions. An initial example revealed the transacetalation between a D-threitol derivative and a diacetal to generate a very complex mixture of cyclic and linear compounds.[17] Acetal exchange between formaldehyde acetals to explore cyclophane formation via DCC has also been presented,[18] while more recently the production of a cyclic polyether DCL was reported resulting in the amplification of small macrocyclic library members by ammonium ions.[19]

Beyond the use of reversible reactions involving carbonyl compounds, numerous other covalent bonds have been utilized in the context of DCLs, most of which provide symmetrical linkages as opposed to the unsymmetrical covalent bonds discussed previously. Disulfide exchange is arguably one of the most widespread and promising reversible reactions in the generation of DCLs, due to not only its simplicity and robustness, but it is also one of the few reactions that is compatible with most biomolecules. Sanders and coworkers reported the first example of a large dynamic library of macrocycle disulfides in water formed from a range of aliphatic dithiol building blocks, including carbohydrate and amino acid derivatives.[20] This reaction is particularly relevant to the work in this thesis due to its biocompatibility and will be discussed in greater length in chapter two. Both olefin metathesis and the Diels-Alder reaction are included on a very short list of exchange reactions involving carbon-carbon bonds. Lehn and coworkers have demonstrated a Diels-Alder reaction that is designed to be reversible under particularly mild conditions.[21] With the development of efficient catalysts such as Grubbs catalyst, olefin metathesis has become a promising reaction for the preparation of

DCLs. Nicolaou and coworkers used reversible olefin metathesis to prepare a library of vancomycin dimers in water, although not initially intended as a DCC experiment per se. A significant catalyst dependence has also been demonstrated in the templated synthesis of porphyrin boxes from alkene functionalized porphyrin building blocks.[22]

**C. Building block design**

For the successful use of DCC, particularly in the event of a selection process, there are various elements that must be considered in the design and choice of building blocks. Specific functionality is required to facilitate an identified reversible reaction, and the reactive functionality can be incorporated either once or twice in each monomer, depending on whether a library of linear dimeric species or cyclic oligomers are desired. The presence of two reactive groups is generally advantageous in that it results in an increase in the structural diversity of the library, which is particularly crucial in the search for molecular receptors. The generation of species of varying size is feasible, albeit often at an entropic cost for the larger oligomers, however the placement of the reactive groups within each monomer is quite important in dictating the formation of linear versus cyclic oligomers.

The structural diversity desired is often tied to the extent of flexibility incorporated into each designed monomer. When monomers are quite rigid they have a tendency to self-sort, particularly when they differ from each other significantly in terms of size and "bite angle" (Figure 1.3).[1a] In this case, heterodimers tend to be destabilized through ring strain and therefore represent only minor species within the mixture. A degree of flexibility within each monomer can allow for a library of increased diversity, even in the

investigation of a single building block library. Furthermore, diversity can be increased by the incorporation of a small, flexible unit, which can bridge monomers that would otherwise not cyclize or be quite strained.

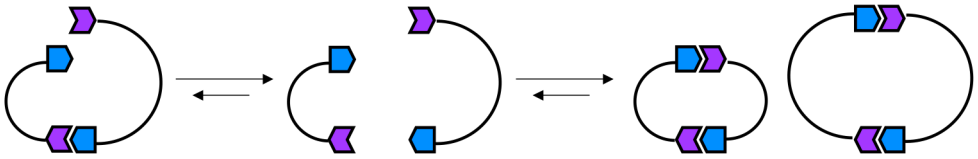


**Figure 1.3.** The tendency of rigid building blocks to self-sort.

A consideration of solubility and molecular weight are also worth nothing in the design of monomers. While often difficult to predict, all species formed in a particular library must maintain complete solubility throughout library equilibration to ensure a complete analysis. This becomes particularly problematic upon the generation of larger library members, as the appropriate solubility of each monomer does not ensure that each macrocycle formed will remain soluble. The solubility must also be maintained in a solvent system that is appropriate for each application, be it templation or otherwise. In the case of templation studies, the hosts and guests must sustain complete solubility as individual components and upon binding. In regards to molecular weight, a unique mass is desired for each monomer, as most libraries are analyzed by mass spectrometry which can not differentiate between structural differences. However, in some cases this can be circumvented as a result of different UV absorption spectra of mass degenerate monomers.

**D. Experimental considerations**

As in most reactions, upon choosing reagents and conditions under which the reaction is performed, the main consideration that remains is the concentration of the reagents. In the case of DCC, the building block concentrations tend to be quite influential on the composition of the resulting library. This is a key parameter because it influences the rates of the exchange reactions, as well as the oligomer to polymer ratio. It has been proposed previously that there exists a critical concentration below which the equilibrium composition of the system consists entirely of small macrocycles.[23] This concentration dependence should be evaluated in the context of each library in order to generate macrocycles of appropriate size and diversity.

In addition, in the pursuit of molecular recognition studies, the template concentration is also crucial, particularly in relation to that of the building blocks. While the most ideal concentrations are often dependent of the level of affinity for a particular template, initial screens are often conducted at template concentrations equal to or above that of the building blocks in order to enforce binding. In cases where the enhancement of a multitude receptors is observed, lowering the template concentration should serve to amplify only those best receptors. Furthermore, theoretical studies have reported the possible breakdown of the correlation between amplification and binding efficiency at certain concentrations as a result of the tendency of a DCL to maximize the binding interactions of the entire library.[24] When choosing between the amplification of a larger number of small, moderate binders versus a smaller number of larger, tighter binders, the latter option is not necessarily preferred, often causing the best binders to go unnoticed.[25] However, this is generally observed when an excess of template is used, and dropping the

template concentration ensures amplification of those best receptors without a serious loss in templating efficiency. This behavior reflects the truly complex nature of such DCL molecular networks.

**E. Possible selection methods**

While initial investigations in this field were focused on library development and the underlying reversible chemistry, attention has more recently moved toward evaluating shifts in product distribution resulting from some type of favorable noncovalent interaction. It is this unique ability of reversible DCLs to efficiently select and amplify a species that is most "fit" in the context of the library that highlights the true utility of this methodology. The majority of the selection methods that have been presented to date involve the use of templating to select for molecules that act as hosts or receptors for particular guests of interest. While synthetic chemists have proven to be quite good at designing elegant syntheses for complex molecular receptors, developing new and modified versions of such receptors often involves a lengthy process of design, synthesis, evaluation, and often redesign. By simply designing potential receptor fragments and leaving the rest up to the recognition properties of the guest, DCC has allowed for the identification and synthesis of quite complex and unpredictable synthetic receptors, with often impressive affinity. This type of selection has been applied to the identification of receptors for a wide range of molecules, from biomolecules to metal ions, using a variety of reversible chemistries and reaction conditions.

Similarly, the selection of new guests or ligands is possible through the introduction of a separate host molecule. This is most applicable in the field of drug discovery, where

the use of an enzyme's active site, for example, can lead to amplification of potential inhibitors. The subtle mechanisms of protein folding, the difficulty in predicting the strength of ligand-protein interactions, and the fact that the structure of many proteins is not precisely known often hampers the design and discovery of new inhibitors by traditional methods. DCC is particularly attractive in this context in that the exact nature of the binding site need not be well understood. Although the association between the ligand and biomolecule is generally entirely noncovalent, a tethering approach has also been developed in which a reversible covalent linkage is used to tether the ligand to the biomolecule.[26] This concept is advantageous in that it facilitates the binding of ligands so that low-affinity binders that would otherwise not bind strongly enough can be detected. It should be noted that such systems can be somewhat limiting due to the physiological conditions and high protein concentrations required.

In addition, the selection of the most stable structure within a mixture of structures with different conformational properties can also be facilitated via DCC. Through the use of reversible chemistry, folding into secondary and tertiary structures by proteins or nucleic acids, as well as polymers or oligomers, can be monitored. Molecular recognition takes place *intramolecularly*, and there is generally no need for a formal template molecule. Somewhat related is the use of DCC to investigate the stabilization of specific library members through *intermolecular* noncovalent interactions. In this case the library composition is biased toward those species that form stable assemblies or aggregates. While this selection method has yet to gain a significant amount of attention, it has much potential in the investigation of self-assembling molecules, such as interlocked or dendritic molecules, or new materials.

## F. Analytical methodology

One primary advantage to DCC as a combinatorial technique is the simultaneous nature of the synthesis and selection processes. Library analysis via liquid chromatography coupled with mass spectrometry has by far evolved as the primary method of library analysis, allowing for a full identification of all species in solution, in both templated and untemplated reactions. Often more important than characterizing all members of the library, LC-MS allows for the accurate identification of all species enhanced due to favorable molecular recognition events (Figure 1.4). The untemplated reaction ideally serves as an internal control for self-selection processes, and allows for a direct comparison to a library containing a molecular guest of interest. Those species enhanced become visually apparent via inspection of the chromatography traces, and subsequent MS analysis facilitates their immediate identification. Further MS-MS can also be utilized in the analysis of libraries involving sequence isomers and regioisomers.[27]

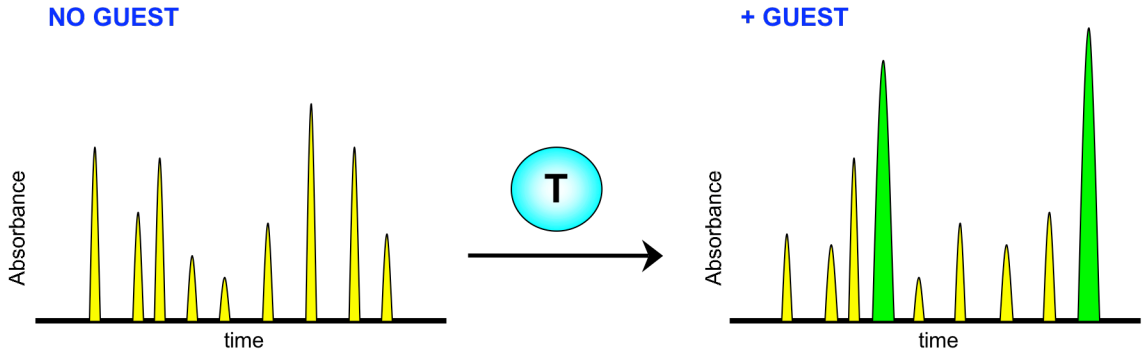



**Figure 1.4.** Representative chromatography traces of an untemplated library distribution (left), and a library templated with an external template (right). Molecular recognition events guide the amplification of those most favorable receptors (shown in green).

While this is quite an effective and efficient method of analysis, particularly with the development of new analytical systems which allow for more high throughput analysis, it

also presents many challenges, as library mixtures can only become so complex before accurate analysis is no longer possible. When the chromatography becomes quite involved, not only is the library composition difficult to ascertain, but the ability to identify amplified, active compounds is often lost. With increasing complexity the probability of generating strong binders increases, yet this results in concentrations of individual library members that are quite low, often below a reasonable concentration for either detection or amplification. The question of how many compounds a DCL can contain while still allowing the amplification of strong binders to useful concentrations is still under debate, and theoretical studies have aided in this regard.[28] This again highlights the importance of careful monomer design, as maximum diversity is desired within the limits of the analytical method. The use of larger solution phase libraries (> 9000 compounds) have been reported in the successful identification of molecular receptors, however most studies continue to make use of relatively simple libraries.[29]

Due to the intrinsic limitations associated with LC-MS analysis, efforts have been put forth in the development of alternative analytical techniques, the primary goal of which is generally to separate those selected receptors from a complex mixture of species. One way in which this has been achieved is through the immobilization of the template during the equilibration processes, allowing for the simultaneous selection, amplification, and isolation of favorable receptors, as the selected compounds remain bound to the resin. Immobilized receptors, such as the carbohydrate-binding protein concanavalin A, have been reported,[30] as well as immobilized guest molecules, such as methylated ammonium ions[31] and cinchona alkaloids.[32] In addition, Miller and coworkers have developed the technique of resin-bound DCC (RBDCC) in which the library constituents themselves are

immobilized, and the detection of selected hits is facilitated by the use of fluorescently labeled targets.[33] Beyond liquid-solid phase segregation, the use of liquid-liquid phase segregation has also be reported using two immiscible solvents to increase the diversity and scope of libraries that can be generated and explored via DCC.[34] Lastly, as an extension to traditional DCL selection methods, Rayner and coworkers have described a process that combines DCC and SELEX (for Systematic Evolution of Ligands by Exponential enrichment) for the *in vitro* selection of small molecule conjugated RNA aptamers that bind tightly to the TAR (transactivation-responsive) element of HIV-1.[35]

**G. Purpose of this work**

In recent years, many contributions have been made to the field of DCC and it has found application among a variety of disciplines, ranging from drug discovery to material science.[36] Although the potential utility of DCC is unquestionable, it remains a relatively limited and young field, allowing for further contributions in a range of areas, while a number of challenges have yet to be addressed. Although the groundwork has been laid in the development of reversible reactions for library generation, new types of reactions will continually be of interest, while the use of known reactions under more physiological-like conditions is undoubtedly desired for biological applications. The potential to generate new monomers and libraries with novel functionality is unlimited, while there are also numerous targets, biological and otherwise, for which DCC may provide an effective method in the identification of host molecules.

This thesis represents an effort to contribute to the field of DCC in a meaningful and diverse way. It is our goal to design novel monomers with potentially useful

functionality, generate new types of libraries for future applications, and to investigate the recognition of biologically interesting targets. The identification of synthetic receptors for post-translationally modified amino acids such as methylated lysine and arginine, as well as methylated nucleotides, has been investigated due to the critical role of each in controlling gene expression. Our interests in both photochemistry and the development of libraries containing more than one reversible reaction prompted an investigation of azobenzene containing monomers. An array of thioester peptide monomers were also evaluated for their ability to readily generate libraries of cyclic macyrocycles for future DCC assays, while the application of larger peptide monomers with known secondary structures have also been investigated in the context of dynamic combinatorial libraries. It is our hope that these studies ultimately play a role in expanding the use and application of DCC within the scientific community, while also highlighting DCC as a versatile tool in the drug discovery process, particularly in the field of epigenetics.

[1] For reviews, see: (a) Corbett, P. T.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J-L.; Sanders, J. K. M.; Otto, S. *Chem. Rev.* **2006**, *106*, 3652-3711. (b) Lehn, J-M. *Chem. Eur. J.* **1999**, *5*, 2455-2463. (c) Ladame, S. *Org. Biomol. Chem.* **2008**, *6*, 219-226. (d) Ludlow, R. F.; Otto, S. *Chem. Soc. Rev.* **2008**, *37*, 101-108. (e) Otto, S. *Curr. Opin. Drug Disc. Devel.* **2003**, *6*, 509-520.

[2] For example, see: (a) Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *Science* **2002**, *297*, 590-593. (b) Baolu, S.; Stevenson, R.; Campopiano, D. J.; Greaney, M. F. *J. Am. Chem. Soc.* **2006**, *128*, 8459-8467.

[3] (a) Lam, R. T. S.; Belenguer, A.; Roberts, S. L.; Naumann, C.; Jarrosson, T.; Otto, S.; Sanders, J. K. M. *Science* **2005**, *308*, 667-669. (b) Hamilton, D. G.; Feeder, N.; Teat, S. J.; Sanders, J. K. M. *New J. Chem.* **1998**, 1019-1021.

[4] For example, see: (a) Crego Calama, M.; Hulst, R.; Fokkens, R.; Nibbering, N. M. M.; Timmerman, P.; Reinhoudt, D. N. *Chem. Commun.* **1998**, 1021-1022. (b) Hof, F.; Nuckolls, C.; Rebek, J. Jr. *J. Am. Chem. Soc.* **2000**, *122*, 4251-4252.

[5] For example, see: (a) Baxter, P. N. W.; Khoury, R. G.; Lehn, J.-M.; Baum, G.; Fenske, D. *Chem. Eur. J.* **2000**, *6*, 4140-4148. (b) Mal, P.; Schultz, D.; Beyeh, K.; Rissanen, K.; Nitschke, J. R. *Angew. Chem. Int. Ed.* **2008**, *47*, 8297-8301. (c) Albrecht, M.; Blau, O.; Frohlich, R. *Chem. Eur. J.* **1999**, *5*, 48-56.

[6] Brady, P. A.; Sanders, J. K. M. *J. Chem. Soc. Perkin I* **1997**, 3237-3253.

[7] Swann, P. G.; Casanova, R. A.; Desai, A.; Frauenhoff, M. M.; Urbancic, M.; Slomczynska, U.; Hopfinger, A. J.; LeBreton, G. C.; Venton, D. L. *Biopolymers* **1996**, *40*, 617.

[8] (a) Eldred, S. E.; Stone, D. A.; Gellman, S. H.; Stahl, S. S. *J. Am. Chem. Soc.* **2003**, *125*, 3422-3423. (b) Hoerter, J. M.; Otte, K. M.; Gellman, S. H.; Cui, Q.; Stahl, S. S. *J. Am. Chem. Soc.* **2008**, *130*, 647-654.

[9] Larsson, R.; Ramstrom, O. *Eur. J. Org. Chem.* **2006,** *2006*, 285-291.

[10] Woll, M. G.; Gellman, S. H. *J. Am. Chem. Soc.* **2004**, *126*, 11172-11174.

[11] Huc, I.; Lehn, J. M. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 2106-2110.

[12] (a) Polyakov, V. A.; Nelen, M. I.; Nazarpack-Kandlousy, N.; Ryabov, A. D.; Eliseev, A. V. *J. Phys. Org. Chem.* **1999**, *12*, 357-363. (b) Nazarpack-Kandlousy, N.; Zweigenbaum, J.; Henion, J.; Eliseev, A. V. *J. Comb. Chem.* **1999**, *1*, 199-206.

[13] Cousins, G. R. L.; Poulsen, S. A.; Sanders, J. K. M. *Chem. Commun.* **1999,** 1575-76.

[14] For example, see: (a) Lam, R. T. S.; Belenguer, A.; Roberts, S. L.; Naumann, C.; Jarrosson, T.; Otto, S.; Sanders, J. K. M. *Science* **2005**, *308*, 667-669. (b) Furlan, R. L. E.; Yiu-Fai, N.; Cousins, G. R. L.; Redman, J. E.; Sanders, J. K. M. *Tetrahedron* **2002**, *58*, 771-778.

[15] Furlan, R. L. E.; Ng, Y.-F.; Otto, S.; Sanders, J. K. M. *J. Am. Chem. Soc.* **2001**, *123*, 8876.

[16] Voshell, S. M.; Gagné, M. R. *J. Am. Chem. Soc.* **2006**, *128*, 12422-12423.

[17] Fuchs, B.: Nelson, A.; Star, A.; Stoddart, J. F.; Viidal, S. *Angew. Chem. Int. Ed.* **2003**, *42*, 4220-4224.

[18] Cacciapaglia, R.; Di Stefano, S.; Mandolini, L. *J. Am. Chem. Soc.* **2005**, *127*, 13666-13671.

[19] Berkovich-Berger, D.; Lemcoff, N. G. *Chem. Commun.* **2008**, 1686-1688.

[20] Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *J. Am. Chem. Soc.* **2000**, *122*, 12063-12064.

[21] Boul, P. J.; Reutenauer, P.; Lehn, J.-M. *Org. Lett.* **2005**, *7*, 15-18.

[22] Van Gerven, P. C. M.; Elemans, J. A. A. W.; Gerritsen, J. W.; Speller, S.; Nolte, R. J. M.; Rowan, A. E. *Chem. Commun.* **2005**, 3535-35537.

[23] Jacobsen, H.; Stockmayer, W. H. *J. Chem. Phys.* **1950**, *18*, 1600-1607.

[24] (a) Grote, Z.; Scopelliti, R.; Severin, K. *Angew. Chem. Int. Ed.* **2003**, *42*, 3821-3825. (b) Severin, K. *Chem. Eur. J.* **2004**, *10*, 2565-2580. (c) Corbett, P. T.; Otto, S.; Sanders, J. K. M. *Chem. Eur. J.* **2004**, *10*, 3139.

[25] Corbett, P. T.; Sanders, J. K. M.; Otto, S. *J. Am. Chem. Soc.* **2005**, *127*, 9390-9392.

[26] Erlanson, D. A.; Hansen, S. K. *Curr. Opin. Chem. Biol.* **2004**, *8*, 399.

[27] Nazarpack-Kandlousy N.; Nelen, M. I.; Goral, V.; Eliseev, A. V. *J. Org. Chem.* **2002**, *67*, 59-65.

[28] Corbett, P. T.; Otto, S.; Sanders, J. K. M. *Org. Lett.* **2004**, *6*, 1825-1827.

[29] Ludlow, R. F.; Otto, S. *J. Am. Chem. Soc.* **2008**, *130*, 12218-12219.

[30] Ramstrom, O.; Lehn, J.-M. *ChemBioChem* **2000**, *1*, 41-48.

[31] Roberts, S. L.; Furlan, R. L. E.; Cousins, G. R. L.; Sanders, J. K. M. *Chem. Commun.* **2002**, 938-939.

[32] Besenius, P.; Cormack, P. A. G.; Liu, J.; Otto, S.; Sanders, J. K. M.; Sherrington, D. C. *Chem. Eur. J.* **2008**, *14*, 9006-9019.

[33] McNaughton, B. R.; Miller, B. L. *Org. Lett.* **2006**, *8*, 1803-1806.

[34] (a) Pérez-Fernández, R.; Pittelkow, M.; Belenguer, A. M.; Sanders, J. K. M. *Chem. Commun.* **2008**, 1738-1740. (b) Choudhary, S.; Morrow, J. R. *Angew. Chem. Int. Ed.* **2002**, *41*, 4096-4098.

[35] Bugaut, A.; Toulmé, J.-J.; Rayner, B. *Org. Biomol. Chem.* **2006**, *4*, 4082-4088.

[36] (a) Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *Drug Discov. Today* **2002**, *7*, 117-125. (b) Karan, C.; Miller, B. L. *Drug Discov. Today* **2000**, *5*, 67-75. (c) Lehn, J-M. *Prog. Polym. Sci.* **2005**, *30*, 814-831.

# CHAPTER II

# SMALL MOLECULE RECEPTORS FOR PROTEIN POST-TRANSLATIONAL MODIFICATIONS

(Reproduced, in part, with permission from The Royal Society of Chemistry and Ingerman, L. A.; Cuellar, M. E.; Waters, M. L. *Chem. Commun*. **2010**, *46*, 1839-1841.)

## A. Background and Significance

### i. Histone post-translational modifications and their role in gene transcription.

Developing a better understanding of the factors that control gene expression is crucial to the study of both development and disease, and consequently, epigenetics has emerged as a major challenge and field of study.[1] In studying the control of genetic information, it has become clear that post-translational modifications (PTMs), particularly those of histone proteins, are intimately involved in gene regulation. These modifications encompass a range of structural changes to natural amino acids, including methylation, acylation, phosphorylation, ubiquitination, and glycosylation, among others, and they have equally diverse functions. Post-translational modifications have been seen to play a critical role in the regulation of signaling pathways and in tagging proteins for degradation, while they can also act as chemical switches to induce or repress protein-protein interactions.[2]

Histone proteins are largely considered to be the framework for packaging genetic material. Within the nucleus, DNA is wrapped around histone proteins to form nucleosomes, which are then further condensed to give chromatin (Figure 2.1). Chromatin can exist as tightly packed inactive genetic material, known as heterochromatin, or switch to actively transcribed genetic material in a lightly packed state, or euchromatin.

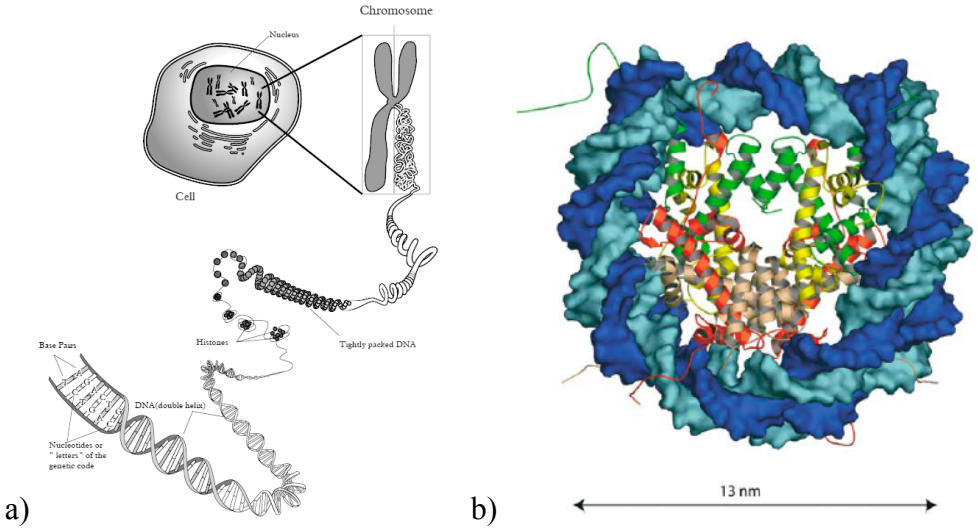



**Figure 2.1.** a) Packaging of DNA within the nucleus (Adapted from a National Human Genome Research Institute Image). b) The atomic structure of the nucleosome core (pdb 1eqz), with each strand of DNA shown in a different shade of blue, surrounding the histone octamer.

There are eight histone proteins which assemble to form an octamer, composed of two H2A-H2B heterodimers and two H3-H4 heterodimers.[3] While DNA is wrapped around the core of this octamer, each histone protein contains disordered termini which protrude from the nucleosome. By extending beyond the nucleosome, these unstructured peptide tails are readily available for enzymatic modification (Figure 2.2). Lysine and serine residues in particular are abundant throughout the histone tail sequences, resulting in highly methylated, acylated, and phosphorylated proteins.

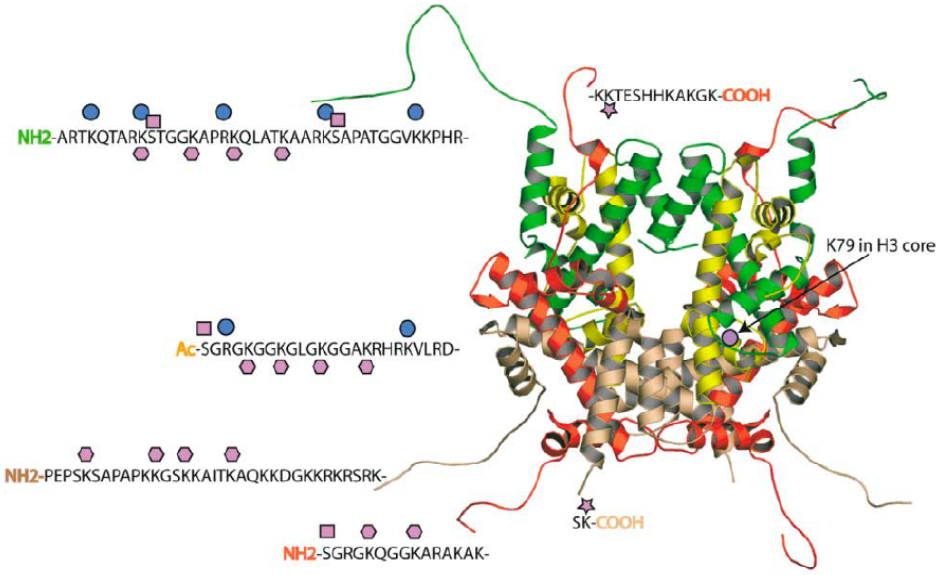



**Figure 2.2.** The types of post-translational modifications observed on the core histones, comprised of a tetramer of H3 (green) and H4 (yellow) and two dimers of H2A (red) and H2B (beige). The modifications are shown, including methylated lysine and arginine (blue circle), acetylated lysine (pink hexagon), phosphorylated serine (pink square), and ubiquitinated lysine (pink star).[4]

While it was initially proposed that these histone modifications exerted their biological function by significantly altering the nucleosome structure, it has since been realized that they instead act as markers for the initiation or repression of specific interactions or for the precise recruitment of non-histone proteins, dictating the higher-order chromatin structure in which DNA is packaged and resulting in gene expression or gene silencing depending on the type and location of the PTM.[5] While lysine acylation is commonly associated with transcriptional activation, as it is thought to negate the favorable electrostatic interaction between lysine and the phosphate backbone of DNA resulting in the loosening of chromatin, lysine methylation often initiates the recruitment of other proteins through specific binding interactions with the modified tail, resulting in further chromatin condensation.[6] Lysine methylation is especially important for chromatin function due to its stability and direct contribution to heritable patterns of gene

expression. Although various PTMs have been documented for over forty years, it has only been within the past decade or so that the impact of certain modifications on gene transcription have emerged.

Solving the puzzle of histone PTMs is complicated by the fact that the same chemical modification occurring at different positions within a single protein can have different effects on gene expression. This is particularly apparent with methylation. For example, within the histone 3 tail, methylation of lysine 4, 36, or 79 results in activation of transcription, whereas methylation of lysine 9 or 27 results in transcriptional repression. This situation is further complicated by the fact that both arginine and lysine can be variably methylated. Histone methyltransferases can add up to three methyl groups to a single lysine side chain, while arginine can exist as either monomethyl or dimethyl arginine (Figure 2.3). Some enzymes specifically produce symmetric dimethylarginine (sRMe2), while others produce asymmetric dimethylarginine (aRMe2). It has become clear that *both* the position of the residue and the degree of methylation together dictate the transcriptional outcome of the modification.[7]

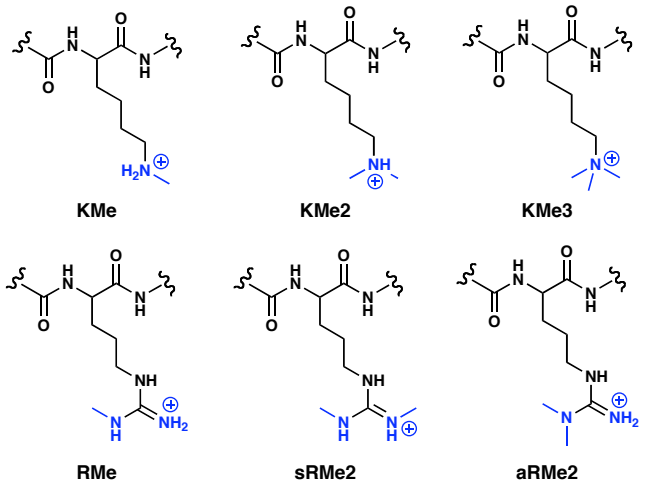



**Figure 2.3.** Methylated lysine and arginine PTMs.

**ii. The histone code.** With the investigation of these modifications and their specific functions in isolation, it soon became apparent that there is a form of communication between PTMs which allows them to work together in a very elaborate signaling pathway. This idea that distinct histone modifications, on one or more tails, act sequentially or in combination to form a code that is read by other proteins to bring about distinct downstream events has since been deemed the "histone code."[8] Thus, it is considered to be the effect of multiple modifications that together control gene transcription.

Mechanistically it can be envision that this communication between modifications may occur at several different levels. First, the presence of multiple modifications (particularly those resulting in an increase or decrease of charge such as phosphorylation and acetylation, respectively) are thought to amplify the readout of signaling pathways, causing greater changes in the overall charge density of histone tails, and leading to a greater change in the extent of chromatin condensation (or decondensation). Second, it is possible that the covalent modification of a histone tail by one enzyme influences the rate or efficiency with which a second enzyme follows. Many sites of modification are close enough within the histone tail to influence, either positively or negatively, the ability of enzymes to further modify the protein. An enzyme may recognize its substrate more effectively in the context of other modifications, while the catalytic activity of an enzyme may also be compromised by prior modification of its substrate. Thirdly, the recruitment of PTM binding proteins may be affected by the presence of multiple modifications, in some cases creating a stronger association, and in others working in tandem to recruit unique biological complexes. PTMs have also been identified which work together as

switches to control gene transcription, whereby adjacent PTMs establish a switch mechanism to regulate protein-protein interactions. One such example is seen with phosphorylation of serine 10 of the H3 protein, which disrupts the binding interaction between the neighboring methylated lysine 9 and heterochromatin protein 1 (HP1) (Figure 2.4).[9] By ejecting the HP1 protein, this PTM switch serves to reverse the effects of histone methylation and in turn reactive transcription. Furthermore, the reversibility of these modifications allows for an extremely intricate control of gene transcription.

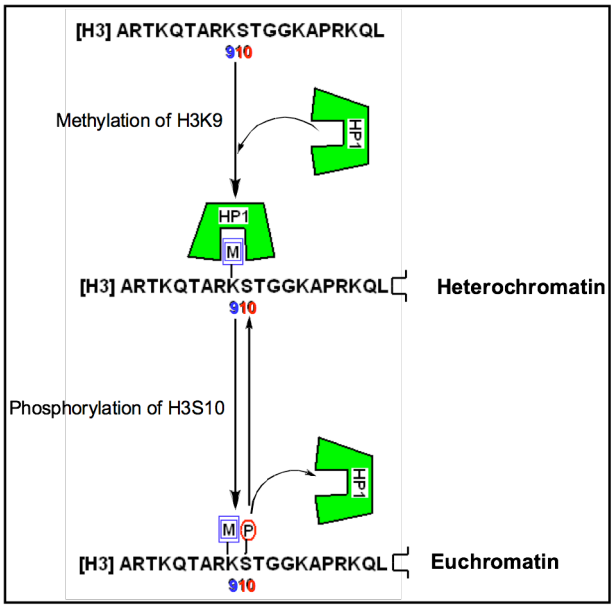



**Figure 2.4.** Phosphorylation as a PTM switch in the H3 tail, serving to reverse the transcriptional outcome of lysine methylation.

**iii. Post-translational modification binding domains.** Although the mechanisms by which cells decipher a PTM-mediated histone code are far from understood, the emerging school of thought is that histone PTMs are read by protein receptors, or effector proteins, which facilitate downstream events via the recruitment and/or stabilization of chromatin-templated machinery.[10] Many epigenetic readers have been identified that bind to specific PTMs of interest, and these have played a key role in furthering our

understanding of how histone PTMs regulate biological functions.[11] These conserved protein domains bind histone PTMs in a way that is dependent on both the type of modification and the position within the histone sequence. The identification of binding modules for methylated lysines has been largely successful, and are known to include chromodomains, Tudor domains, MBT-repeats, WD40-repeats, and PHD fingers, whereas protein receptors for methylated arginines in histone proteins have yet to be identified.[12] The identification of such methylated arginine reader proteins remains a challenge, as does the broader goal of uncovering the relationship between PTM binding proteins and human disease.

Unlike the elimination of charge upon lysine acetylation, all methylated forms of lysine are cationic at physiological pH, with trimethyllysine containing a fixed positive charge. As the number of lysine methyl groups increase, the hydrophobicity and the distribution of positive charge of the methylammonium group increases, while concomitantly it's ability to serve as a hydrogen bond donor decreases. As a result, each PTM requires a protein partner that can adapt to these inherent physical properties, in turn leading to a great deal of specificity.

The recognition of these modifications results largely from contacts made between the methylammonium group and aromatic residues in the protein receptor, forming an aromatic "cage" about the PTM (Figure 2.5). These aromatic cages are generally highly specific for a certain methylation state, discriminating between PTMs based on only very slight differences in size and shape. The cages are thought to be preorganized and static in nature, showing little or no appreciable structural perturbation upon binding, and therefore minimal loss in entropy. The binding interaction between the

methylammonium and the aromatic cage is mediated largely by cation-π interactions, while hydrophobic desolvation effects also have an appreciable role. The cation-π interaction is generally thought of as a charge-quadrupole interaction between a positively charged species and an aromatic ring, primarily electrostatic in nature. More specifically, as the quadrupole moment places partial negative charge above each face of the aromatic ring, favorable interactions with cations occur perpendicular to the plane of the ring. The role of this interaction in proteins is a topic of great interest, particularly in the investigation of protein structures, protein-protein interactions, and protein-ligand interactions.[13]

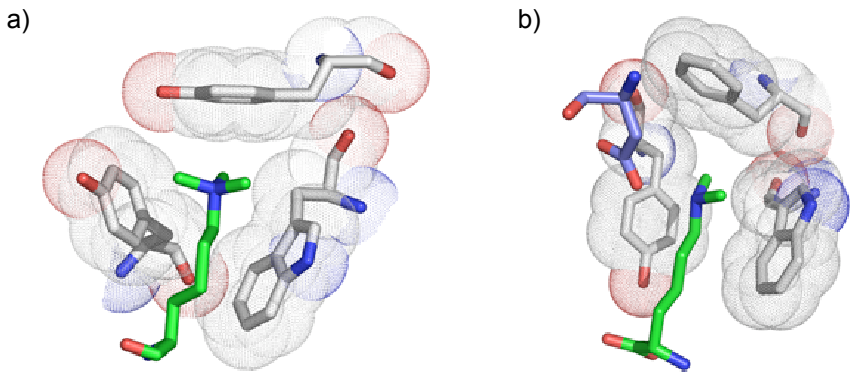



**Figure 2.5.** Binding pockets for lysine PTMs. a) Histone 3 K9Me3 (green) bound to the HP1 chromodomain (gray) via cation-π interactions with three aromatic side chains (pdb 1KNE). B) Histone 3 K4Me2 (green) bound to the 3-MBT domain (gray) via cation-π interactions with three aromatic residues and a hydrogen bond to glutamic acid (blue) (pdb 2RHI).

The magnitude of the cation-π interaction in proteins is dependent on numerous factors, including the electron density of the aromatic ring (for example, phenylalanine versus tryptophan), the distribution of positive charge across the cation, the degree of solvent exposure of the interaction, as well as the contribution of other forces such as van der Waals and hydrophobic interactions. Prior work in the Waters Lab has shown the

importance of the charge-quadrupole interaction to KMe3 binding and specificity in the context of a β-hairpin model system, as well as in a histone peptide, indicating that the driving force for binding is indeed the cation-π interaction as opposed to the hydrophobic effect.[14] Specifically, replacement of KMe3 in a histone 3 peptide with its neutral analogue, tert-butylnorleucine, resulted in binding to the HP1 chromodomain nearly as weak as unmethylated lysine, demonstrating the essential nature of the cation-π interaction to the recognition of the H3 tail by the HP1 chromodomain with good affinity and selectivity.

In the recognition of the lower methylation states, hydrogen bonding and steric exclusion also become increasingly important. Nearby acidic residues in the protein are also known to form salt bridges with the protonated amines of KMe2 and KMe. This has been demonstrated in the context of a PHD domain by engineering dimethyllysine recognition specificity. While the wild-type PHD finger's binding preference is for histone 3 K4Me3 over K4Me2, the binding preference was reversed for that of KMe2 over KMe3 through the mutation of a key tyrosine residue to glutamic acid.[15] The change in selectivity is associated with hydrogen bonding between the KMe2 proton and the carboxylate group of the glutamic acid side chain.

Furthermore, complexes with a lower methylation state PTM are bound via the cavity-insertion recognition mode, where the methylammonium group is inserted into and buried deep within a protein cleft. This allows the protein pocket to bind only those PTMs of the appropriate size.[12] In contrast, the higher methylation states are known to use a surface-groove recognition mode whereby the binding pockets are both wider and more accessible. The methyllysine side chain lies along the protein surface groove and

consequently the effector proteins have slightly less stringent preferences for specific methylation states.

Beyond selectivity for different methylation states, effector proteins display marked selectivity for the sequence context within which a PTM is presented. The PTM containing peptides are generally unstructured in the unbound state, but undergo an induced-fit conformational change, adopting secondary structure when bound to their recognition proteins. The histone tail peptides tend to adopt a β-sheet conformation, pairing through an antiparallel alignment with an exposed face of an existing β-sheet of the effector protein, as seen in the case of H3 K4Me3 recognition by the PHD finger of NURF.[16] Similarly, six residues of the H3 peptide containing K9Me3 insert as a β-strand on the surface of the protein receptor, HP1 chromodomain, and complete a β-sandwich overall fold.[17] While backbone hydrogen bonds between residues adjacent to the PTM and the effector protein are formed, strengthening the interaction, the sequence selectivity is largely due to complementary side chain interactions between the two domains. The steric compatibility, intermolecular hydrogen-bonding, and electrostatic interactions of the surrounding residues all play a role in determining the binding affinity and specificity.

A significant protein-protein interaction induced by lysine methylation most relevant to the work presented in this thesis is the binding of histone 3 K9Me3 to the HP1 chromodomain, which results in gene silencing (Figure 2.6). It is thought that the function of the chromatin binding is to passively stabilize a dense conformation of the chromatin fiber associated with gene repression.[18] The binding affinities have been reported to be in the micromolar range for both H3 K9Me3 and H3 K9Me2, ranging from 2.5 to 10 μM for K9Me3 and from 7 to 15 μM for K9Me2, while there was no observable

binding for the corresponding unmethylated peptide.[14,17] The chromodomain adopts an incomplete β-barrrel secondary structure, with the methyllysine binding pocket positioned at one end. The aromatic binding cage is made up of three aromatic residues, two tyrosines and a tryptophan, to form a conserved aromatic pocket into which the methylammonium group inserts itself. Mutation of any of these aromatic residues drastically reduces affinity for the methylated histone tail. Furthermore, residues 5-10 of the histone tail ($QTARK_9S$) interact with the chromodomain by an induced-fit sandwiching between terminal β-strands, completing a five-stranded antiparallel β-sheet. Mutation studies of residues in both the peptide and protein have confirmed the contribution of intermolecular contacts along the extended surface groove to both binding affinity and selectivity.[19]

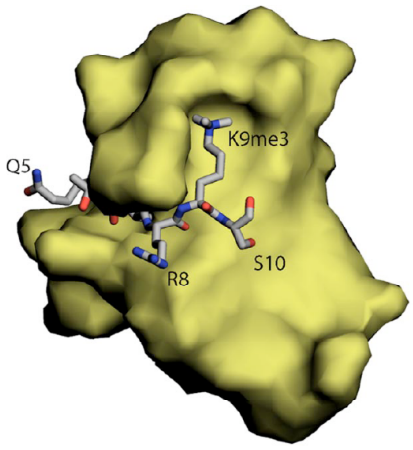



**Figure 2.6.** Crystal Structure of the HP1 chromodomain (yellow surface) in complex with Lys9Me3 H3 tail residues 5 through 10 (gray stick).

To discuss all of the remaining lysine methyl binding domains would be exhaustive and beyond the scope of this thesis, however there are two general classes of protein folds that bind methyllysine with aromatic cages, sharing several molecular recognition features: members of the Royal superfamily of folds (chromodomains, Tudor domains,

MBT domains, and others) and PHD fingers. While some proteins, such as the HP1 chromodomain or the Polycomb (PC) protein, contain a chromodomain that allows them to specifically recognize the appropriate repressive methylation mark (H3 K9 and H3 K27 respectively), others such as the chromodomain helicase DNA-binding protein 1 (CHD1) uses its chromodomain to bind the activating methylated H3 K4.[20] Therefore, the ultimate function of the methylated lysine PTM is a reflection of the type of effector protein it has evolved to interact with in a highly specific manner, either an activator or a repressor of transcription.

While much attention has been focused on the methylation of lysines, arginine methylation has similarly been identified as a key player in the regulation of cellular processes.[21] While knowledge of methylarginine effector proteins is presently limited, as no structure of a reader bound to this PTM is available, there is much evidence for methylated arginine acting as a mediator of protein-protein interactions.[22] Similarly to methylated lysine, it is expected that the recognition of methylated arginine may often depend on its interaction with aromatic rings. There is much precedence for the stacking of arginine with aromatic residues, as cation-π interactions are generally observed between the guanidine sidechain and tyrosine, tryptophan, and phenylalanine.[23] Arginine's ability to stack with aromatic residues, even more so than lysine, is due to it's ability to interact via a combination of cation-π and π-π stacking interactions, lending itself to both favorable dispersion and electrostatic forces without a significant desolvation penalty. Methylation of arginine therefore is expected to magnify this Arg-π interaction, with a further increase in hydrophobicity. Prior work in the Waters lab has investigated the interaction between tryptophan and aRMe2 and sRMe2 in the context of

a β-hairpin peptide, showing that methylation of arginine significantly enhances peptide stability while maintaining the stacked geometry.[24] This suggests that this enhanced interaction may be a key driving force in the mediation of protein-protein interactions.

There is substantial evidence that the SMN Tudor domain, which is linked to spinal muscular atrophy, recognizes the arginine-glycine rich C-terminal tails of spliceosomal Sm proteins, and that this binding is mediated by symmetrical dimethylation of arginine side chains.[25] It is proposed that this interaction is facilitated by positioning the symmetrical dimethylarginine side chain near a cluster of conserved aromatic residues, forming a typical cage-like mode of recognition. Although knowledge of methylarginine effector proteins is limited, some arginine residues have been identified that influence other PTMs and other PTM-protein interactions. The methylation of histone 4 Arg3 has been seen to facilitate histone 4 acetylation, enhancing the activation of transcription by nuclear hormone receptors.[26] Reports have also suggested that asymmetric dimethylation of arginine at histone 3 Arg2 antagonizes methylation at Lys4, likely due to the fact that an unmodified H3Arg2 is essential for binding of the lysine methyltransferase, and effective binding of the methyltransferase is required for H3 Lys4 methylation.[27] It is clear that like other PTMs, methylated arginine residues also participate in a form of crosstalk with other modifications, determining the final outcome of recognition events and in turn regulating the complicated cascade controlling gene transcription.

**iv. Current methods for identifying post-translational modifications.** Although much progress has been made, efforts to map out histone modifications, as well as study their functional interplay and subsequent biological outcomes has to some degree been limited by the analytical techniques available to do so. First, antibodies have been used

in a variety of analyses, particularly Western blots, which have evolved as a widespread and easy method of analyzing protein modifications. Although they can easily be used on large mixtures of proteins, antibody-based analyses are met with their own set of caveats. Due to the sequence specificity of antibodies, antibody-based analyses rely on prior knowledge of the type and position of specific protein modifications, as well as the availability of high-quality antibodies.[28] This clearly prevents the use of antibodies in the discovery of novel sites of modification. Furthermore, issues of antibody production, unwanted cross-reactivity, epitope occlusion, and the inability to detect patterns of modifications to investigate the histone code also plague such antibody-based analyses.[29]

Mass spectrometry has become a versatile and indispensable tool in proteomics, allowing for determination of the mass-to-charge ratio of protein ions, while tandem MS (MS/MS) enables smaller peptides to be sequenced and PTMs to be identified and characterized. Some of the challenges associated with MS based analyses include the mass degeneracy of specific modifications, the low occupancy at many sites (for example, the presence of only 5% of modified protein may suffice to activate a signaling pathway), the reversibility of some modifications, and the difficulty of purification and sample preparation.[30] Despite these limitations, many histone PTMs have been identified by mass spectrometry, and new details of the histone code are continually being reported due to the effectiveness of MS analysis.[31]

Most common is a "bottom-up" type of MS analysis which involves the digestion of histone proteins into short peptides, enabling sequencing analysis and characterization of the individual fragments (Figure 2.7b). While this method allows for determination of the exact location of the PTM, it becomes difficult to determine which peptides

originated from the same molecule in a complex mixture, and information about interdependence of modifications is lost.[32] In contrast, "top-down" analysis allows for the identification of the number of modifications with an intact protein, without providing information on the location of such modifications (Figure 2.7a). Several more recent techniques such as electron capture dissociation have proved particularly useful, especially in the context of histone PTMs, allowing for the sequencing of intact proteins.[33] Kelleher and coworkers, among others, have used this technique quite successfully in the characterization of histone PTMs on all four core human histones.[34]

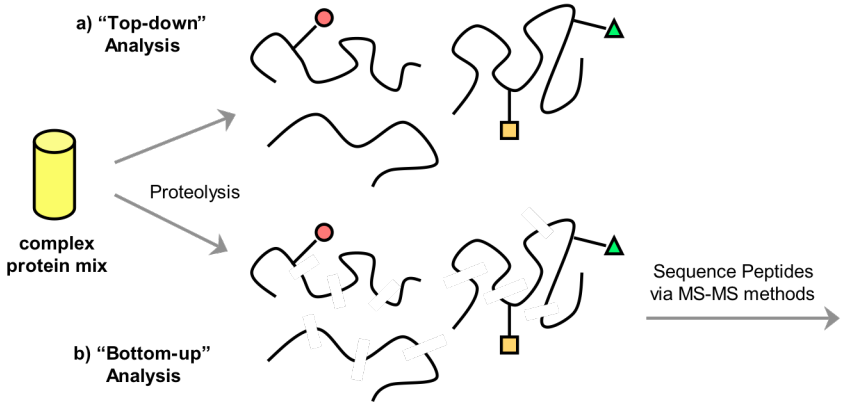



**Figure 2.7.** a) "Top-down" and b) "bottom-up" approaches to identifying PTMs via mass spectrometry.

Recently, DNA aptamers have also been used to develop protein affinity reagents for acylated lysine PTMs, as aptamers have received much attention as possible alternatives to traditional antibodies.[35] Aptamers are pieces of single-stranded DNA or RNA that fold into three-dimensional structures with binding sites that are complementary in shape and charge to the target antigens, and because these molecules can be produced *in vitro*, their recognition and binding properties can be tailored to specific targets of interest.

Although aptamers serve as a new molecular tool for the recognition of PTMs, like antibodies, they require a prior knowledge of the target and are sequence dependant.

**v. Goal of this work.** The development of small molecules which mimic PTM effector proteins is a field which up to this point has remained largely unexplored, as utilizing synthetic receptors for molecular recognition often presents the challenge of poor selectivity. Furthermore, synthetic modification for optimization is often quite difficult and labor intensive. However, synthetic receptors have some significant potential advantages, including better reproducibility, lower cost, lower molecular weight, and the possibility of being used within cells. Small molecule receptors are also much more likely to bind PTMs in a sequence independent fashion due to their size, unlike most known protein receptors. The only report of a synthetic molecule that distinguishes between post-translational modifications was presented very recently, making use of the well-known ammonium ion binder *p*-sulfonatocalix[4]arene to selectively recognize trimethyllysine.[36]

Given the importance of mapping out PTMs in attempt to unravel the histone code and better understand their pivotal responsibility in controlling gene expression, non-sequence selective small molecule receptors have the potential to serve as novel reagents in the discovery of *new* sites of methylation within proteins. The work described in this chapter uses dynamic combinatorial chemistry to provide a high-throughput method for the identification of small molecule receptors for methylated lysines and arginines. While we are not looking to achieve selectivity for a PTM solely within the context of its native sequence, identifying receptors which can distinguish between the different lysine and arginine methylation states is of great importance. DCC allows for screening of the

same library against all lysine and arginine methylation states, providing a rapid approach to screen for selectivity for the different PTMs. We report here the identification of multiple synthetic receptors selective for various methylated PTMs, one of which recognizes trimethyllysine with both comparable affinity and selectivity as the native HP1 chromodomain.

**B. Library design and precedence for using dynamic combinatorial chemistry for the recognition of ammonium cations.**

Dynamic combinatorial chemistry is an attractive alternative to the rational design of synthetic receptors in that it allows molecular recognition to guide the synthesis of complex host systems from simple building blocks.[37] These building blocks are linked reversibly under thermodynamic control to produce an equilibrium mixture of potential receptors. In the presence of a molecular target, favorable host-guest binding interactions drive the synthesis and amplification of the best receptor(s) at the expense of other oligomers.

Disulfide exchange has become one of the most widely used reactions in dynamic combinatorial libraries, and it is particularly suited for this biological application as it occurs in aqueous solution at close to neutral pH. Disulfide exchange is advantageous in that disulfides form readily from thiols in the presence of oxygen, it takes place under mildly basic conditions in the presence of a catalytic amount of thiol, it can be quenched under acidic conditions, and disulfides are stable toward many different functional groups.[38] During oxidation the mixture contains both disulfides and thiols, allowing for equilibration through nucleophilic attack of thiolate anions on the disulfides, displacing a

new thiolate anion in the process (Figure 2.8). The primary downside to disulfide exchange is that it occurs relatively slowly, sometimes requiring on the order of weeks for equilibration to be reached.

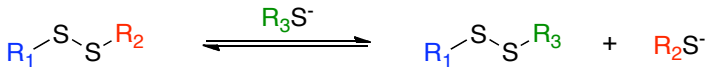



**Figure 2.8.** Reversible disulfide exchange mediated by thiolate anions.

In choosing a set of dithiol monomers for this application, there were various requirements to consider. First, the facile synthetic introduction of the thiol groups is required to facilitate reversible disulfide exchange. Second, in attempt to mimic the recognition of methylated PTMs by aromatic cages in protein receptors, aromatic surfaces are needed to allow for potential cation-π and hydrophobic interactions. Third, functional groups such as carboxylates are required to induce solubility in aqueous solutions, while carboxylates also have the potential to induce favorable electrostatic interactions. Last, a mixture of structural monomers, some more curved and/or rigid and others more linear and/or flexible, are necessary to avoid either self-sorting or extensive linear oligomerzaiton of monomers.

Building blocks were utilized that contain structural elements of previously reported synthetic receptors for specific guests, equipped with appropriate thiol functionality for reversible covalent attachment. The water soluble cyclophane receptor utilized by Dougherty and coworkers (Figure 2.9) has been demonstrated as an effective cation binder.[39] Taking advantage of this known receptor, Sanders and coworkers have made use of subunits **A** and **B**, which allow for the introduction of two thiol groups on each with minimal effects on the overall structure of the receptor. The cyclophane has also

been shown to be tolerant of minor structural variations such as the nature of the phenyl spacer, providing evidence that replacing the ether linkage with a disulfide would not necessarily destroy binding.[40]

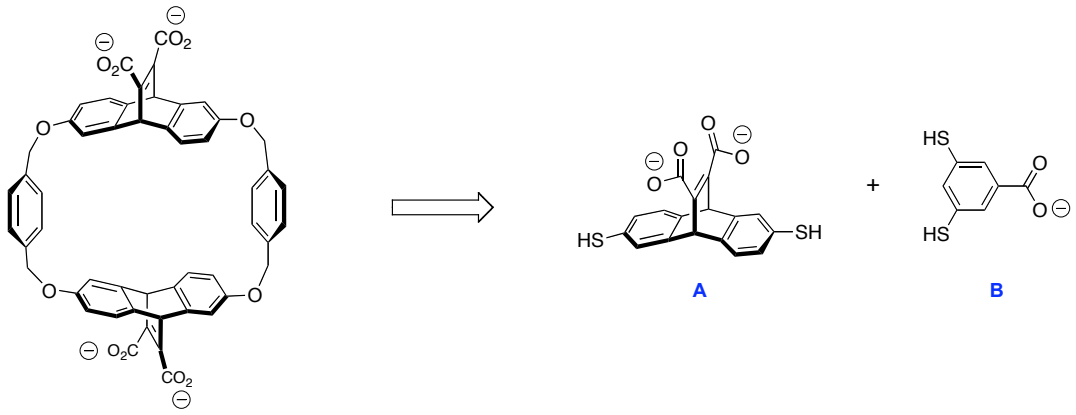



**Figure 2.9.** Design of the dithiol building blocks **A** and **B** inspired by Dougherty's cyclophane receptors.

Sanders and coworkers have elegantly demonstrated the use of such dithiol monomers in isolation as single component libraries, or in combination, developing libraries of higher complexity. In initial reports investigating the exposure of such libraries to cationic guest molecules, it was seen that receptors $\mathbf{A_2B}$ and $\mathbf{A_3}$ were significantly amplified in the presence of methylisoquinoline and *N*-methyl morphine respectively (Figure 2.10).[41] Methylisoquinoline showed about a four-fold selectivity for $\mathbf{A_2B}$ over $\mathbf{A_3}$ (5 μM versus 22 μM), whereas *N*-methyl morphine showed a twenty-fold selectivity for $\mathbf{A_3}$ over $\mathbf{A_2B}$ (39 μM versus 1.9 μM). Interestingly, the structures of the selected receptors differ from the expected tetrameric disulfide analogue of Dougherty's cyclophane receptor. This may be due to the more flexible nature of the ether linkages as compared to their corresponding disulfides, while the increased length and different bond angles of the disulfides may cause a difference in binding. It is also feasible that the

smaller binding pockets of the selected disulfide receptors provide higher affinity for the ammonium guests.

A later report demonstrated the very efficient amplification of tetrameric receptor **$A_4$** by tetramethylammonium in a highly diasteroselective fashion, as dithiol **A** is synthesized and introduced into the library as a racemic mixture.[42] Both the affinity and selectivity were much greater in this case, as tetramethylammonium binds **$A_4$** with a binding constant of 250 nm in comparison to 154 μM for **$A_2B$** and 1.2 mM for **$A_3$**. The higher affinity is attributed to the fact that **$A_4$** can easily fold into a four-stave barrel shape, forming a cavity ideally sized to accommodate tetramethylammonium. It is evident that there are some structural similarities between tetramethylammonium and trimethyllysine, and we envisioned that libraries containing similar building blocks could also be used to effectively develop small molecule receptors for methylated PTMs.

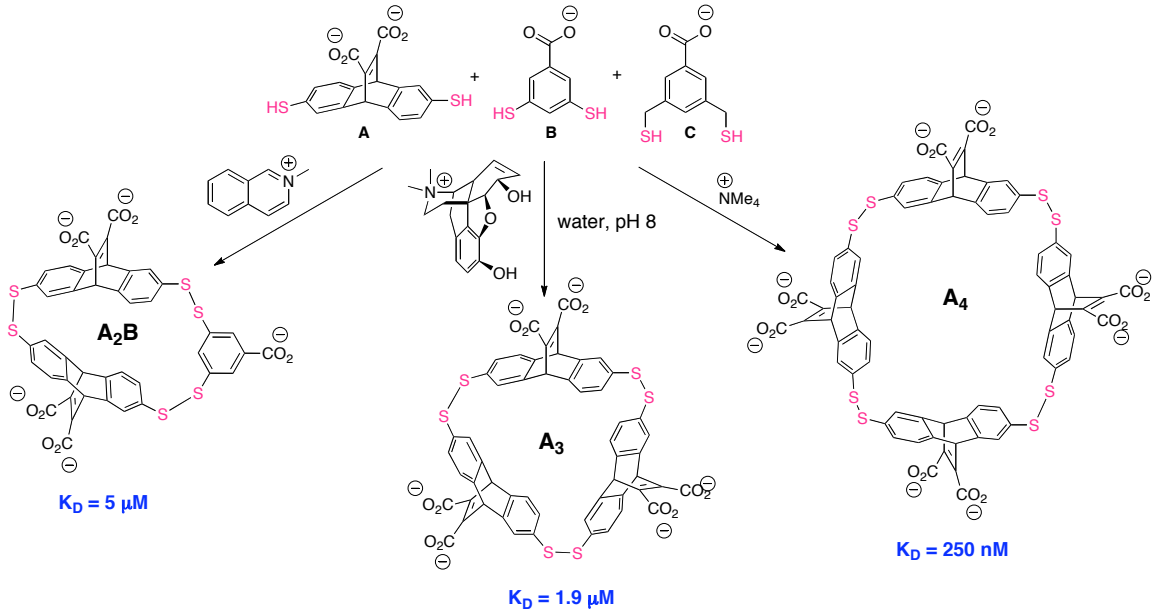



**Figure 2.10.** Selective amplification and binding of three different hosts with three different quaternary ammonium guests.

Showing an even higher degree of structural similarity to trimethyllysine, the biologically relevant template acetylcholine has been more recently investigated in a study using a large series of different but structurally related cationic templates.[43] In this case acetylcholine was seen to bind the corresponding receptors with more modest affinity and selectivity (Figure 2.11), nonetheless we saw this as further evidence that dynamic combinatorial chemistry would serve as an effective method to identify PTM receptors.

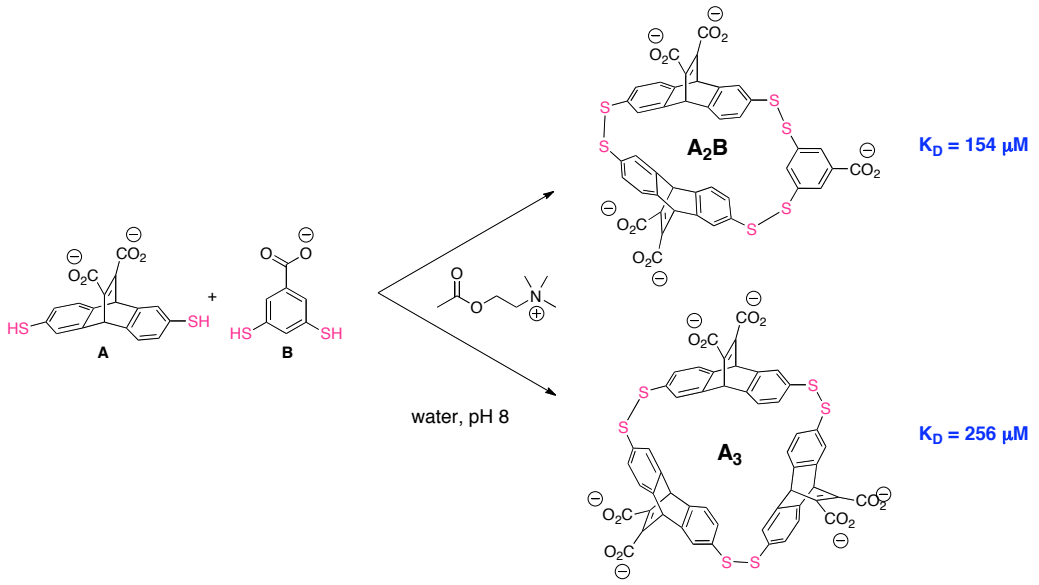



**Figure 2.11.** Macryocyclic receptors $\mathbf{A_2B}$ and $\mathbf{A_3}$ amplified by acetylcholine.

In addition to dithiol monomers **A**, **B**, and **C**, we designed and synthesized a range of building blocks to investigate, all containing the basic features of aromatic groups, carboxylates, and thiols (Figure 2.12). Monomer **H** is the only exception, which we envisioned as a possible linker for any macrocycles that would not otherwise be able to cyclize. The size of the aromatic surface was varied in different monomers, along with the position of and distance between the two thiols, and the position of the carboxylic acid. Variation of the monomers in a fairly systematic manner allows for optimization of

both binding affinity and selectivity, and this is arguably the most notable advantage to developing receptors in this fashion, as such a systematic study using traditional macrocyclic receptors may present various synthetic challenges. Generally three to four monomers were combined in a single library, with the size of the libraries largely limited by the analytical analysis. The majority of the following discussion is focused on those libraries in which the equilibrium was significantly perturbed in the presence of post-translational modifications due to the recognition of the PTMs by specific receptors.

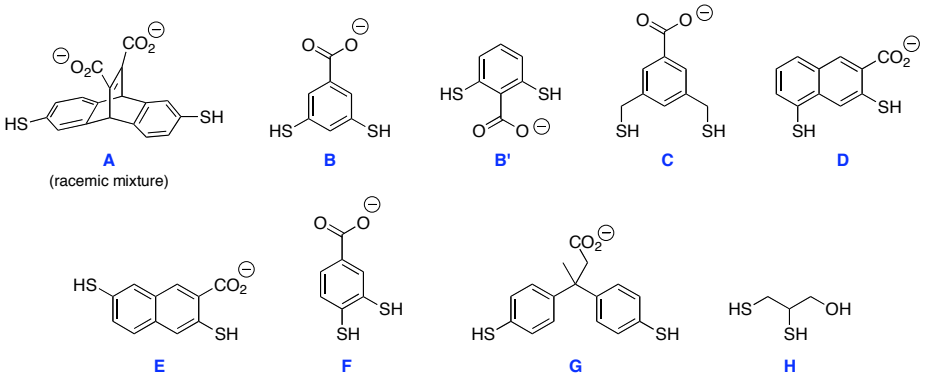
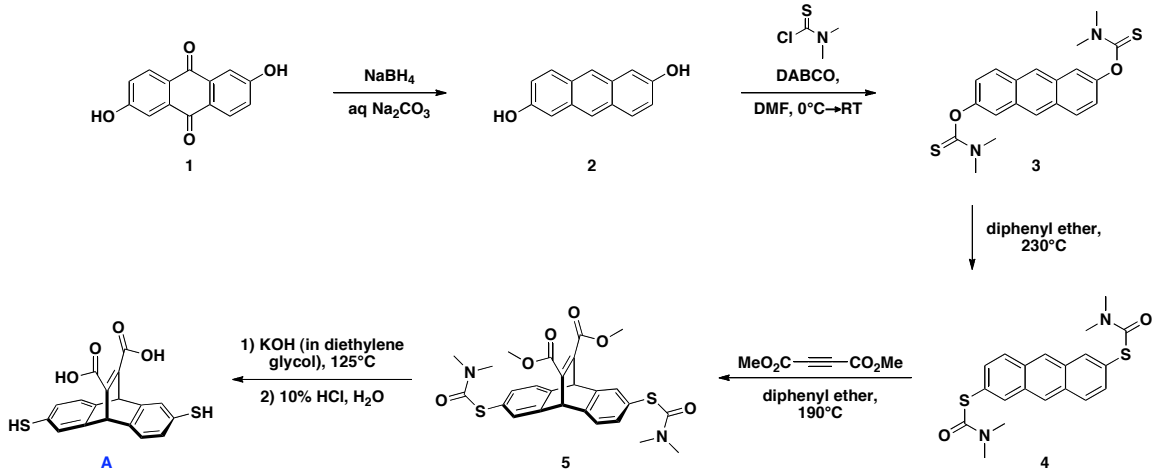



**Figure 2.12.** Monomers used in DCLs.

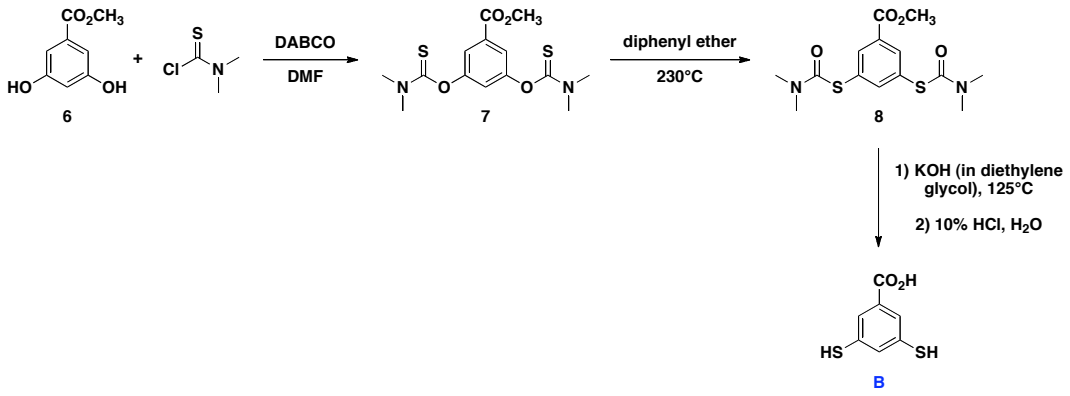
## C. Selective recognition of trimethyllysine: *receptors rac-$A_2B$ and meso-$A_2B$*

### i. Results and discussion

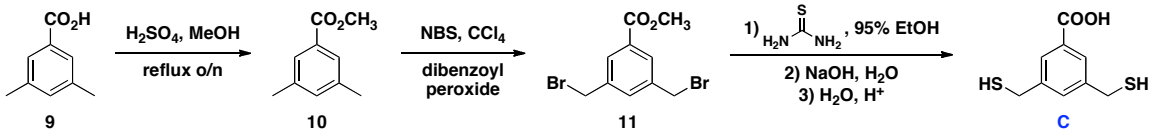
**a. Design and synthesis of guests and monomers.** Based on the successful use of monomers **A**-**C** in the recognition of ammonium cations previously, a library that we sought to investigate early on was that containing those same three monomers. Monomers **A**, **B**, and **C** were synthesized as shown in Schemes 2.1, 2.2, and 2.3 respectively, according to literature procedures, with slight modifications in the case of monomer **C**.[43,44]




**Scheme 2.1.** Synthesis of monomer **A**.




**Scheme 2.2.** Synthesis of monomer **B.**




**Scheme 2.3.** Synthesis of monomer **C**.

Monomer **A** was synthesized as a racemic mixture, where the key step is a Diels-Alder reaction between the protected precursor **4** and dimethylacetylenedicarboxylate. For the majority of the dithiol monomers, including **A** and **B**, starting from aromatic alcohols, the thiol groups are introduced through a Newman-Kwart rearrangement on the

O-thiocarbamate intermediates (**3** and **7**). Monomer **C** is prepared uniquely through a radical bromination followed by introduction of the thiol functionality with thiourea.

Dipeptides were designed as PTM templates for initial library screens (Figure 2.13), each of which contained a C-terminal glycine and a post-translational modification on the N-terminus (Ac-PTM-Gly-$NH_2$). The peptides were also acylated on the N-terminus to avoid any potential non-specific interactions with the free amine. The peptides were synthesized by Fmoc solid phase peptide synthesis, with the LysMe3 guest easily obtained via alkylation of LysMe2 with methyl iodide.[45] Each peptide was characterized by electrospray mass spectrometry.

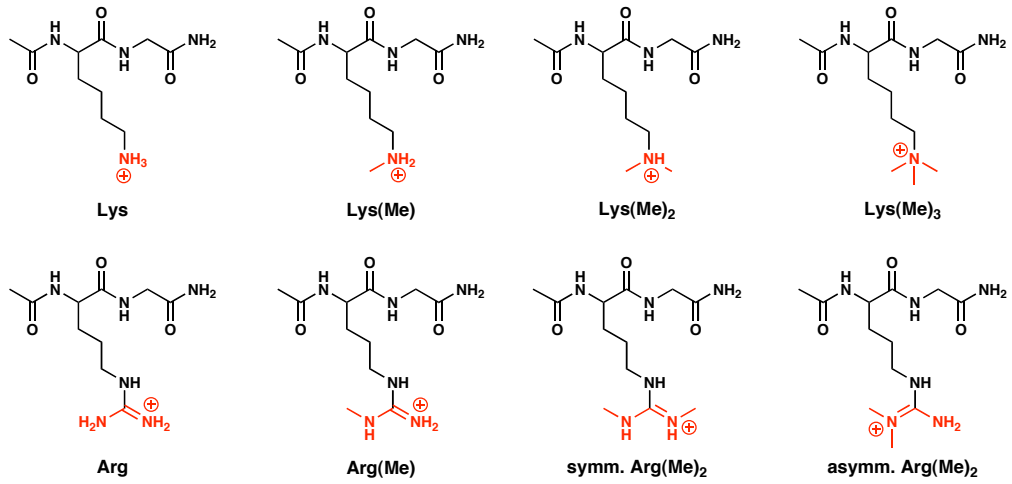



**Figure 2.13.** Methylated and unmethylated lysine and arginine dipeptide guests.

**b. Unbiased library templation studies.** Dynamic combinatorial libraries were prepared by mixing equimolar amounts of monomers **A**-**C** (2.5 mM each) in basic solution. The building blocks were dispersed in water and the pH was adjusted to 8.5 with NaOH and HCl. Eight DCLs contained a PTM template at a concentration of 7.5 mM, while no template was added to the ninth DCL. The libraries were allowed to oxidize in capped HPLC vials at room temperature for about three weeks. All libraries

remained completely soluble with no visible precipitation, which is essential. The library compositions were analyzed by LC-MS at various time points, and all library members could be easily identified by negative electrospray ionization. In some cases, the mass degeneracy of multiple macrocycles prevented the complete characterization of all species without further MS/MS analysis (for example, **$A_2BC$** and **$B_6$** have identical masses). Upon equilibration the chromatograms were integrated, and in each case the peak areas were represented as a percentage of the total chromatogram area. This aided in taking into account slight baseline differences from chromatogram to chromatogram.

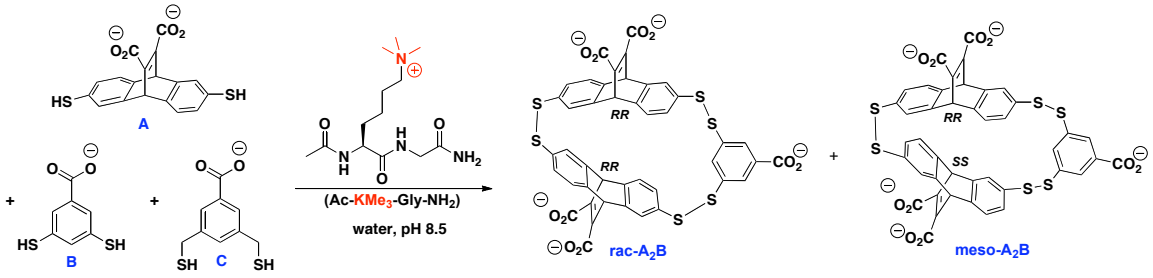



**Figure 2.14.** Dynamic combinatorial library of dithiol monomers resulting in amplification of receptors *rac*-**$A_2B$** and *meso*-**$A_2B$** upon addition of a $KMe_3$ guest.

In the absence of a guest, two major constituents were present, **BC** and **ABC**, as well as numerous smaller peaks corresponding to other macrocyclic library members. However, the introduction of a trimethyllysine dipeptide resulted in a significant and promising shift in the library composition in comparison to the untemplated reaction (Figure 2.15). The amplification of two peaks in the KMe3 templated library were clearly observed, both corresponding to the trimeric macrocycles *rac*- and *meso*-**$A_2B$** (Figure 2.14). The two isomers were found to exist in a ratio of about 5.5:1 in the untemplated library, with *rac*-**$A_2B$** as the larger of the two peaks. The identification of *rac*-**$A_2B$** as the dominant isomer in an untemplated library was previously described by

Sanders and coworkers with the use of a chiral shift reagent to give separate NMR signals for the two enantiomers of the racemate without changing the signals for the *meso* isomer, and confirmed in our case by $^{1}$H NMR upon isolation of the macrocycles.[43]

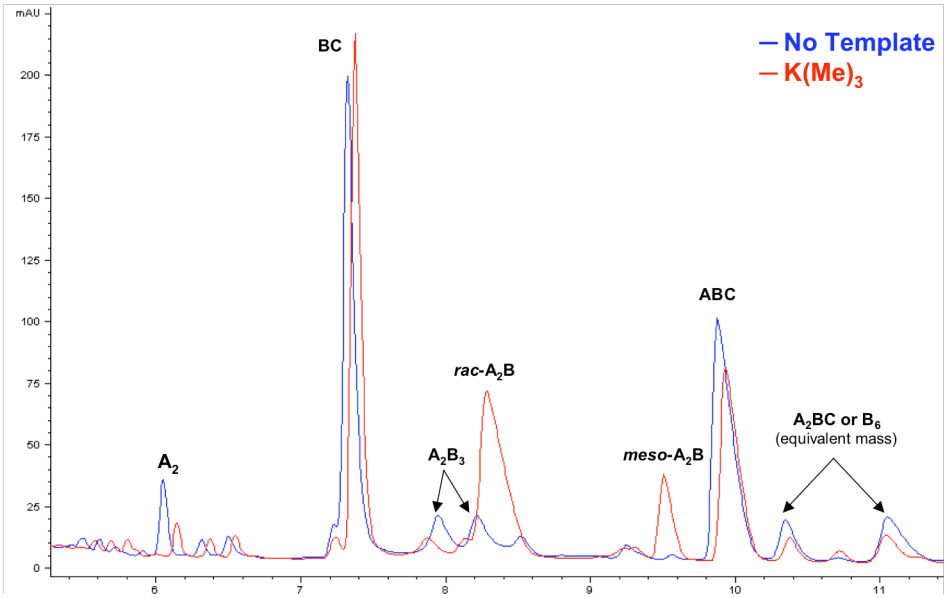



**Figure 2.15.** Part of the analytical HPLC trace at 254 nm of a DCL consisting of monomers **A**, **B**, and **C** (2.5 mM each), untemplated (blue) and in the presence of Ac-$KMe_3$-G-$NH_2$ (red).

Due to our interest in not only identifying a macrocycle that amplifies trimethyllysine, but one that is selective as well, the amplification of $\mathbf{A_2B}$ in the presence of lower lysine methylation states was also examined. It was found that the amplification of both $\mathbf{A_2B}$ diastereomers was dependent on the extent of methylation, with amplification increasing with increasing methylation, suggesting significant selectivity for higher methylation states (Figure 2.16). This trend mimics that of the HP1 chromodomain's affinity for methylated histone tails, displaying the greatest affinity for the trimethylated peptide tail.

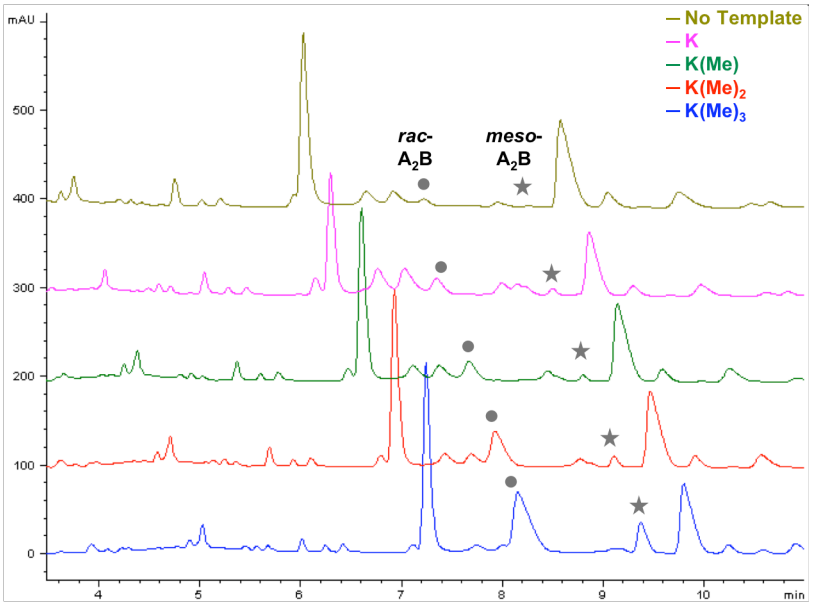



**Figure 2.16.** Analytical HPLC traces at 254 nm of DCLs consisting of monomers **A**, **B**, and **C** (2.5 mM each), untemplated (gold) and in the presence methylated lysine dipeptides (pink = Lys, green = LysMe, red = $LysMe_2$, blue = $LysMe_3$). *Rac*-$\mathbf{A_2B}$ is marked by a grey circle in each, and *meso*-$\mathbf{A_2B}$ by a grey star.

The libraries templated with methylated arginine dipeptides were also investigated to determine whether these PTMs were recognized by the same receptors as the methylated lysines, or whether the structural differences between the two methylated amino acids are significant enough to result in either no amplification or the amplification of different receptors. The chromatograms of the arginine templated libraries revealed amplification of both *rac*- and *meso*-$\mathbf{A_2B}$ receptors, however the extent of amplification is modest in comparison to that of LysMe3 (Figure 2.17). Although it is somewhat surprising that the same macrocycles are amplified, it seems logical that the three methyl groups on lysine would result in a greater number of contacts with the receptor and hence produce greater amplification than that of its dimethylated arginine counterparts.

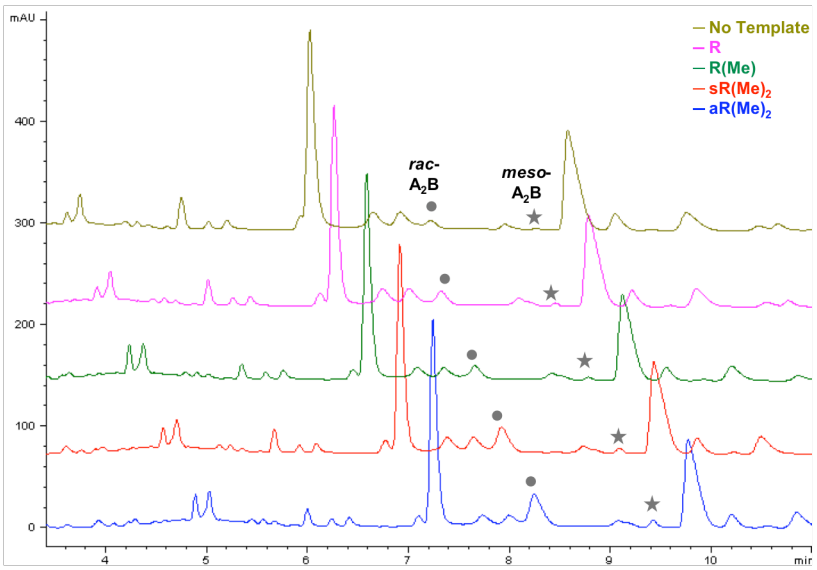



**Figure 2.17.** Analytical HPLC traces at 254 nm of DCLs consisting of monomers **A**, **B**, and **C** (2.5 mM each), untemplated (gold) and in the presence methylated arginine dipeptides (pink = Arg, green = ArgMe, red = symmetric $ArgMe_2$, blue = asymmetric $ArgMe_2$). *Rac*-$\mathbf{A_2B}$ is marked by a grey circle in each, and *meso*-$\mathbf{A_2B}$ by a grey star.

The library chromatograms were integrated after a month of equilibration, and the percent area of each $\mathbf{A_2B}$ peak was calculated by dividing the raw peak area by the total area of the chromatogram, not including the area of the peptide guest peak in the case of the templated reactions. It can be noticed that there is a significant amount of broadening in the $\mathbf{A_2B}$ peaks, and the area of *meso*-$\mathbf{A_2B}$ in the untemplated reaction is quite small making an extremely accurate integration difficult. In that regard it must be noted that the integration values are used to represent an overall trend as opposed to an absolute comparison of binding affinity. The percent change in each peak area upon amplification can be determined with the following equation, as is shown for each case in Figure 2.18.

$$\% \text{ change} = [(\% \text{ area templated} - \% \text{ area untemplated}) / \% \text{ area untemplated}] \times 100$$

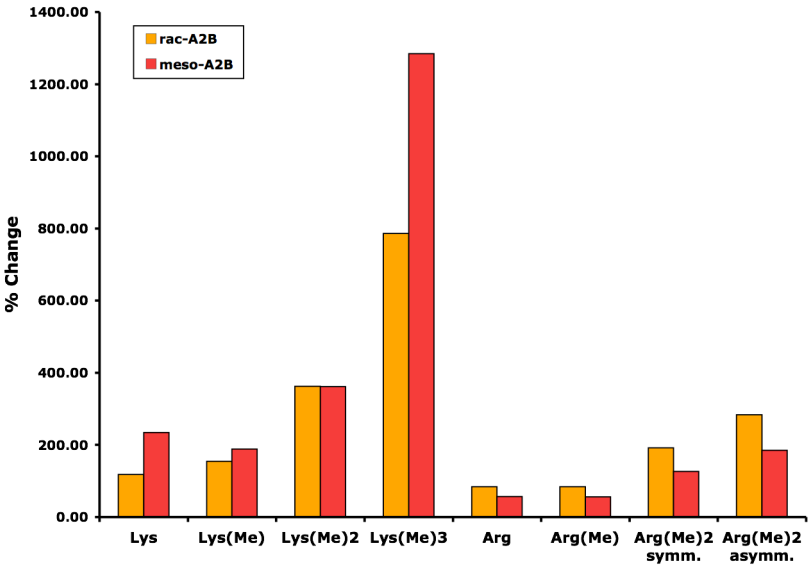



**Figure 2.18.** Amplification results of an **ABC** library templated with all lysine and arginine methylation states, represented as a percent change in amplification. The amplification of *rac*-$\mathbf{A_2B}$ is shown in red, and that of *meso*-$\mathbf{A_2B}$ is shown in orange.

There are likely several factors that result in the observed selectivity of both $\mathbf{A_2B}$ diastereomers for LysMe3 over lower lysine and arginine methylation states. First, the binding pocket may be too large for the lower methylation states, while the increased hydrophobicity and van der Waals interactions of the higher methylation states is more amenable to binding an aromatic pocket. In addition, all other Lys and Arg PTMs can form hydrogen bonds with water, and binding to the pocket may therefore require some degree of desolvation, which would be unfavorable. Model systems have also shown that cation-π interactions are enhanced with increasing methylation of the ammonium group.[46] Thus, KMe3 is expected to have a stronger cation-π interaction with $\mathbf{A_2B}$ than the lower methylation states.

In looking more specifically at the amplification with methylated arginine guests it appears that having two methyl groups in close proximity in the case of asymmetric

dimethyl arginine is more favorable than when the two methyl groups are on different nitrogens in the case of symmetric dimethyl arginine. However, symmetric dimethyl arginine still shows increased amplification in comparison to its monomethylated counterpart indicating that both methyl groups still play a role in binding to **$A_2B$**.

**c. Biased library templation studies.** Focusing on the more promising recognition of trimethyl lysine, an additional DCL was set up biased towards the formation of **$A_2B$** in which building blocks **A** and **B** were mixed in a 2:1 ratio (7.5 mM total) and the selectivity of the diastereomeric receptors for the different methylation states of lysine was investigated. By biasing the library towards the formation of **$A_2B$**, these receptors are formed in a much higher concentration than in the unbiased case. This not only allows for a better visualization of the amplification, but also a more accurate integration of the peak areas.

A diverse library was generated in the absence of a template, with **$A_2$** and **$A_2B_3$** as the dominant species, and *rac*-**$A_2B$** and *meso*-**$A_2B$** constituting only 2.4% and 1.1% of the total library composition respectively. Upon templation with LysMe3, both *rac*-**$A_2B$** and *meso*-**$A_2B$** clearly become the largest species in the library. As expected, the amplification of both diastereomers was dependant on the extent of methylation, suggesting a significant selectivity for the different methylation states, however little diastereoselectivity was observed (Figure 2.19).

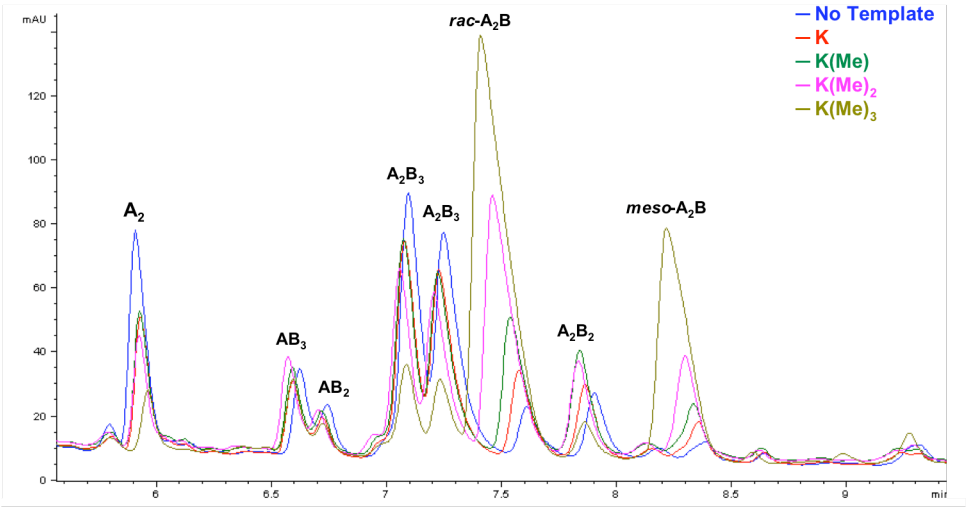



**Figure 2.19.** Part of the analytical HPLC trace at 254 nm of a biased DCL consisting of monomers **A** (5 mM) and **B** (2.5 mM), untemplated (blue) and in the presence of all lysine methylation states, Ac-KMe$_x$-G-NH$_2$.

The amplification was quantified in this case by determining the amplification factor for both **A$_2$B** receptors. Amplification factors are defined as the concentration of a library member in a templated library divided by its concentration in the corresponding untemplated library, and can be calculated by dividing the percent areas of the HPLC peaks corresponding to a specific compound in the templated and untemplated libraries. It must be noted that when no amplification is observed, this results in an amplification factor of 1, not 0. Despite some minor differences, it is evident that the amplification in this biased library follows a similar trend as that of the DCL not biased towards the desired receptor, showing significant selectivity for trimethyllysine over its dimethylated counterpart (Figure 2.20). While the concentration of both **A$_2$B** receptors less than doubles in the presence of unmethylated lysine, it increases by about 1000 percent when templated with LysMe3. These results provided substantial evidence that *rac*-**A$_2$B** and *meso*-**A$_2$B** were worth pursuing as small molecule receptors for trimethyllysine.

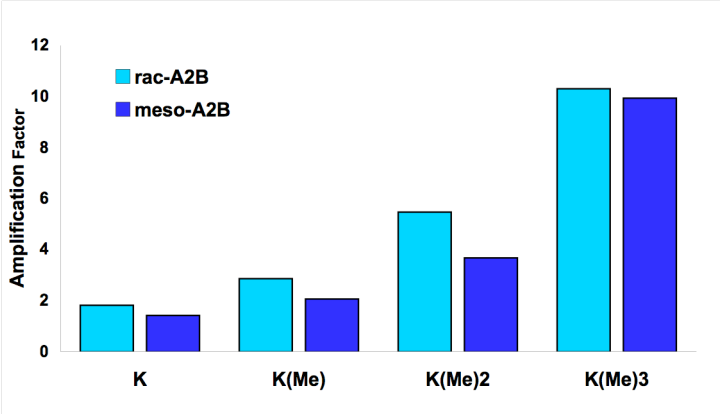
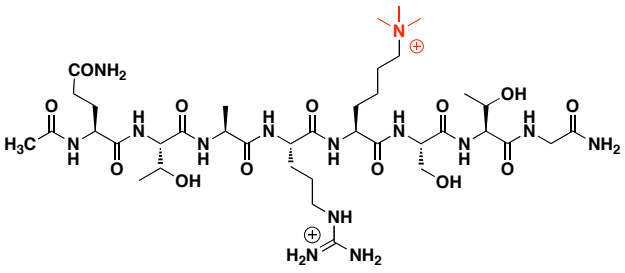
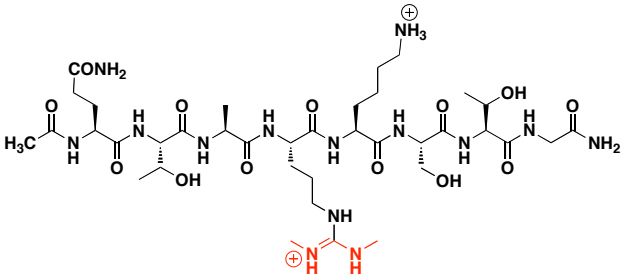



**Figure 2.20.** Extent of amplification of *rac*-$\mathbf{A_2B}$ and *meso*-$\mathbf{A_2B}$ with Ac-K$Me_x$-G-$NH_2$ guests relative to the untemplated library. The amplification of *rac*-$\mathbf{A_2B}$ is shown in light blue, and that of *meso*-$\mathbf{A_2B}$ is shown in dark blue.

**d. Investigation of a histone 3 K9Me3 peptide template.** To ensure that these amplification results were maintained in the context of naturally methylated histone tail peptides, the 8-residue peptides in Figure 2.21 were synthesized. Similar to the dipeptides, they were synthesized by Fmoc solid phase peptide synthesis, with the LysMe3 peptide obtained via alkylation of LysMe2. The sequence contains residues five through twelve of the histone 3 tail, with a lysine at position 9 and an arginine at position 8, both of which can be methylated in nature. Successful amplification of *rac*-$\mathbf{A_2B}$ and *meso*-$\mathbf{A_2B}$ with these longer peptide guests is important to show that binding to the PTM does not occur only in the context of the isolated PTM, and that surrounding residues do not interfere with binding.



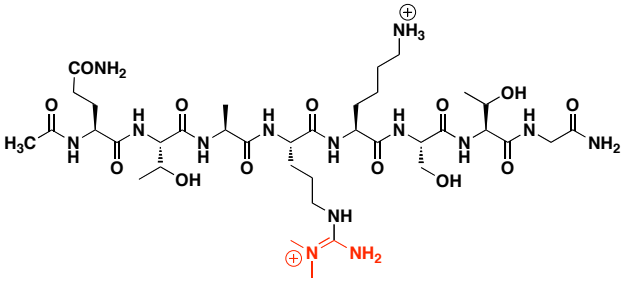




**Figure 2.21.** Histone tail peptide guests containing residues 5-12 from the native histone 3 sequence, and either KMe3 at lysine 9 or RMe2 at arginine 8.

Biased libraries were set up as described previously, monitored by LC/MS, and *rac*-**$A_2B$** and *meso*-**$A_2B$** amplification was quantified (Figure 2.22). It must be noted that over extended periods of time, a small amount of decomposition of these peptides guests was sometimes observed under these basic conditions. Our analysis was performed before this became a problem, yet after equilibration had been reached. Significant amplification of both diastereomeric receptors by H3 K9Me3 was observed, with amplification factors of about 9 for each, close to that of the KMe3 dipeptide. This confirms that **$A_2B$** binding is still possible in the context of a histone tail peptide, greatly

increasing the significance of this work, as many possible applications for such histone tail binding small molecule receptors can be envisioned. In contrast, amplification in the presence of either H3 R8Me2 peptide is minimal, confirming that the previously observed selectivity is maintained. Furthermore, this confirms that **$A_2B$** does not interact significantly with the residues surrounding the PTM to enhance binding beyond what would be expected for the PTM residue alone (for example, it could be predicted that favorable electrostatic interactions would form between the carboxylates on **$A_2B$** and the adjacent basic lysine residue).

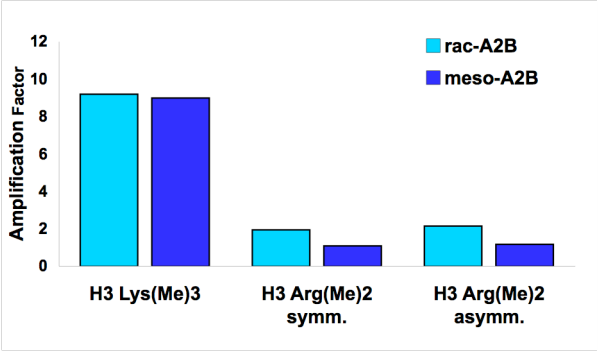



**Figure 2.22.** Extent of amplification of *rac*-**$A_2B$** and *meso*-**$A_2B$** with methylated histone tail guests relative to the untemplated library. The amplification of *rac*-**$A_2B$** is shown in light blue, and that of *meso*-**$A_2B$** is shown in dark blue.

### e. Preparative scale biased libraries and isolation of *rac*-$A_2B$ and *meso*-$A_2B$.

To pursue further investigation of *rac*-**$A_2B$** and *meso*-**$A_2B$** as potential trimethyllysine small molecule receptors, preparation and isolation of these molecules was required. While preparation of the **$A_2B$** diastereomers by traditional synthetic methods would prove daunting, DCC allows for the facile synthesis of these isomers in a fairly straightforward manner. Large scale biased libraries were prepared with a total volume of 10-20 mL and a ratio of **A** : **B** : guest of 2 : 1 : 3, with a total monomer concentration (**A** + **B**) of either 7.5 or 10 mM. Due to the known high affinity of **$A_2B$** for methylisoquinoline ($K_d$ = 5

μM) and our ability to prepare it cheaply on a much larger scale than a KMe3 peptide, most of these libraries were templated with methylisoquinoline to achieve the highest **$A_2B$** concentration possible. The successful amplification of both **$A_2B$** diastereomers by methylisoquinoline to a greater extent than LysMe3 was also confirmed on an analytical scale.

The three solids were combined with the appropriate amount of water, titrated to pH 8.5 with NaOH and HCl, and let stir in a capped scintillation vial for 1-3 weeks before being purified on a semi-preparative HPLC. While separation of the two isomers was easily feasible on an analytical HPLC, co-elution was much more problematic on a semi-preparative system, and separation of the diastereomers was largely facilitated by the use of a column heater. While the dominant *rac*-**$A_2B$** diastereomer was easily purified, a second and sometimes third purification was often required to achieve pure *meso*-**$A_2B$**. The purity of both receptors was confirmed in each case by analytical LC/MS due to their different retention times, as both receptors are indistinguishable by mass. While the receptor synthesis is not very laborious upon successful synthesis of the monomer building blocks, only milligram quantities (generally < 5 mg) are obtained after purification in each case.

**f. NMR analysis of trimethyllysine binding to *rac*-$A_2B$.** To confirm that the binding event is indeed occurring at the site of modification, we chose to investigate the trimethyllysine dipeptide NMR spectra in the absence and presence of *rac*-**$A_2B$**. If *rac*-**$A_2B$** is indeed binding the PTM, an upfield shift of the lysine methyl groups would be expected as a result of being buried in the aromatic pocket of the receptor. The dipeptide

$^1$H NMR spectra appeared as expected, and the chemical shifts of the lysine side chain protons were confirmed by a TOSCY experiment.

Upon introduction of a slight excess of *rac*-**$A_2B$** to ensure the majority of the trimethyllysine is in the bound state, significant broadening of all peptide proton signals was observed, indicating a change in the peptide's environment. Furthermore, the methyl groups were upfield shifted by 0.826 ppm, indicating that the binding event is indeed occurring at the site of modification. When the same experiment was performed at 5°C instead of room temperature, a slightly greater upfield shift of 0.976 ppm was observed for the methyl groups. No NMR studies were performed with *meso*-**$A_2B$** because of limited availability of the compound, largely due to significant co-elution with *rac*-**$A_2B$** during purification, however a similar upfield shift of the lysine methyl group would be expected when bound to *meso*-**$A_2B$**. Overall, these findings confirm that the key binding interaction is between the trimethylammonium group and the aromatic pocket, as in the histone-chromodomain protein-protein interaction.

**g. Binding studies by fluorescence anisotropy.** Fluorescence anisotropy was used to measure the dissociation constant of *rac*-**$A_2B$** and *meso*-**$A_2B$** to each methylated lysine PTM to determine both an absolute binding affinity and the extent of selectivity between the different methylation states. Histone tail peptides were synthesized similar to those used for prior amplification studies, containing residues 5-12 of the histone 3 protein, and each methylation state at Lys9. The peptides were appended with an N-terminal carboxyfluorescein (FAM) for fluorescence detection (Figure 2.23). The binding studies were done at slightly basic pH to ensure complete solubility of **$A_2B$**, as well as to maintain full fluorescein fluorescence, as it is significantly reduced below pH

7.[47] Fluorescein was chosen as the fluorophore over TAMRA (carboxytetramethylrhodamine) in order to minimize any possible binding interactions between **$A_2B$** and the fluorophore, as TAMRA contains two dimethylamine groups appended to the ring system in place of the ketone and hydroxy functionalities of fluorescein.

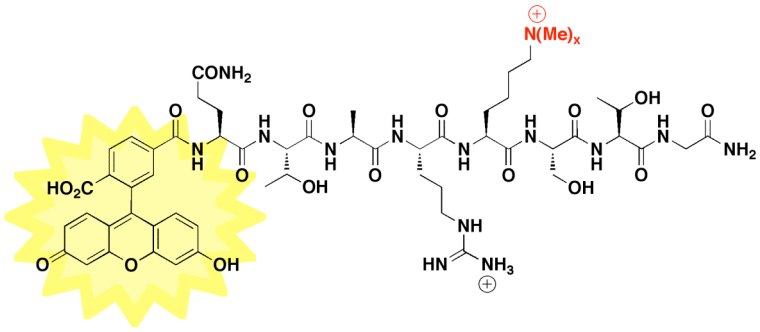



**Figure 2.23.** Fluorescently labeled histone 3 tail peptides for anisotropy experiments, containing and N-terminal fluorescein and a PTM and Lys9 (x = 0, 1, 2, 3).

In each experiment, while the amount of fluorescent peptide remains constant, the concentration of unlabeled receptor concurrently increases. When the fluorophore is excited by polarized light, the loss of polarization in the emitted light can be correlated with the mobility of the fluorophore. Binding of **$A_2B$** to the PTM-containing peptide increases the effective molecular weight of the fluorescent peptide, decreasing the effective fluorophore mobility, and therefore increasing the fluorescence anisotropy. The ratio of parallel and perpendicular polarized light emitted upon binding is used to determine the anisotropy value, which can in turn be correlated to binding affinity.[48] The anisotropy value is determined based on the equation below, where $I_{II}$ and $I_{\perp}$ are fluorescence intensities observed when the emission polarizer is oriented parallel and perpendicular to the direction of the polarized excitation, respectively.

$$r = \frac{I_{II} - I_{\perp}}{I_{II} + 2I_{\perp}}$$

The H3 K9Me3 peptide was found to bind both *rac*- and *meso*-$\mathbf{A_2B}$ with binding affinities of about 25-30 μM, which is virtually equivalent to the binding affinity of the HP1 chromodomain protein receptor (Table 2.1).[14,17,19] This comparable affinity to the chromodomain is quite promising yet somewhat surprising, as $\mathbf{A_2B}$ is significantly smaller in size than the protein receptor, and as a result does not form additional interactions with the adjacent residues of the histone tail. Moreover, upon investigating the binding affinity to the lower lysine methylation states it was determined that H3 K9Me3 binds to both *rac*- and *meso*-$\mathbf{A_2B}$ with greater than 2-fold selectivity over its dimethylated counterpart, exhibiting slightly better selectivity than the native protein. While *rac*-$\mathbf{A_2B}$ bound H3 K9Me2 with a dissociation constant of 58 μM, *meso*-$\mathbf{A_2B}$ bound with a similar affinity of 73 μM. The mono- and umethylated peptides exhibited considerably weaker affinity for *rac*-$\mathbf{A_2B}$, with no appreciable binding for the unmethylated histone tail. The binding affinity of the mono- and umethylated peptides to *meso*-$\mathbf{A_2B}$ was not measured due to a limited amount of *meso*-$\mathbf{A_2B}$, although binding is expected to be weak and comparable to that of *rac*-$\mathbf{A_2B}$. Overall, it is evident that the trend in *both* affinity and selectivity of $\mathbf{A_2B}$ parallels that of the native protein receptor.

**Table 2.1**. Dissociation constants for H3 tail peptides with varying methylation states at K9 determined by fluorescence anisotropy.[a]

| Peptide | *rac*-**$A_2B$** $K_d$ (μM)[b] | *meso*-**$A_2B$** $K_d$ (μM)[b] | HP1 chromodomain $K_d$ (μM) |
|---|---|---|---|
| **H3 K9Me$_3$** | 25 ± 3 | 28 ± 4 | 10[d] (21 ± 2 at 25°C)[e] |
| **H3 K9Me$_2$** | 58 ± 10 | 73 ± 9 | 15[d] (39 ± 7 at 25°C)[e] |
| **H3 K9Me** | 166 ± 50 | N.A.[c] | 96[e] |
| **H3 K9** | > 1200 | N.A.[c] | > 1000[e] |
| **H3 K9Me$_3$ R8G** | 34 ± 8 | N.A.[c] | -- |

[a] Conditions: 27 °C, 10 mM phosphate buffer, pH 8.5. [b] Errors are from the fit. [c] Not available due to limited material (*meso*-**$A_2B$**), although binding is expected to be comparable to that of *rac*-**$A_2B$**. [d] Values are taken from Hughes, R. M.; Wiggins, K. R.; Khorasanizadeh, S.; Waters, M. L. *Proc. Nat. Acad. Sci. USA* **2007**, *104*, 11184-11188; Conditions: 15°C in pH 7.5 phosphate buffer. [e] Conditions: 25°C in 50 mM potassium phosphate buffer, pH 8.0, 25 mM NaCl (Eisert and Waters, *unpublished results*).

To confirm that the adjacent basic arginine residue does not contribute significantly to binding, a mutant H3 K9Me3 peptide tail with glycine in place of Arg8 was synthesized. Fluorescence anisotropy experiments revealed that *rac*-**$A_2B$** bound the R8G mutant with a dissociation constant of 34 μM, which is comparable to that of the native sequence, indicating that Arg8 is not required for binding. This helps to confirm that **$A_2B$** binding to KMe3 is independent of the sequence surrounding the methylated residue. Furthermore, a H3 peptide tail was synthesized where the Arg8 remained intact, however the Lys9 was mutated to glycine. As expected, this mutation resulted in total loss of binding to *rac*-**$A_2B$**, reinforcing that the methylated lysine is required for binding.

To fully evaluate the selectivity of **$A_2B$** for trimethyllysine over other post-translational modifications, fluorescent H3 peptide tails were synthesized with methylated arginine residues at position 8 and unmethylated lysine at position 9. Three

peptides were synthesized, with aRMe2, sRMe2, or RMe in place of R8, and the binding of each to *rac*-**$A_2B$** was similarly evaluated by fluorescence anisotropy (Table 2.2). In each case, binding was determined to be much weaker than that of the methylated lysine peptides, further establishing the selectivity of **$A_2B$** for trimethyllysine. While both aRMe2 and sRMe2 bind with similar weak affinities of about 600 μM and 750 μM respectively, binding to RMe is substantially weaker, comparable to that of the fully unmethylated peptide. The binding affinities for *meso*-**$A_2B$** were not demonstrated, but they are expected to be weak as well. It is clear that in this case, recognition of the methylated guanidinium moiety of arginine is significantly different than that of the methylated ammonium group of lysine and that based on these fluorescence anisotropy binding studies, **$A_2B$** shows promise as a selective receptor for trimethyllysine with affinity paralleling that of the native protein receptor.

**Table 2.2**. Dissociation constants for H3 tail peptides with varying methylation states at R8 determined by fluorescence anisotropy.[a]

| Peptide | *rac*-**$A_2B$** $K_d$ (μM)[b] |
|---|---|
| **H3 asymmetric R8Me$_2$** | 601 ± 234 |
| **H3 symmetric R8Me$_2$** | 754 ± 194 |
| **H3 R8Me** | > 1700 |

[a] Conditions: 27 °C, 10 mM phosphate buffer, pH 8.5. [b] Errors are from the fit.

**h. Binding studies by isothermal titration calorimetry.** To supplement the fluorescence anisotropy binding data, isothermal titration calorimetry (ITC) was used to confirm the binding affinity of *rac*-**$A_2B$** and *meso*-**$A_2B$** to trimethyllysine. Similar results from these two quite different methods of analysis not only serves to confirm our prior

results, but also allows for a more direct comparison to the binding affinities reported previously by Sanders and coworkers, as they used ITC to evaluate binding. In each experiment, a H3 K9Me3 peptide is titrated into a solution of either *rac*-**$A_2B$** or *meso*-**$A_2B$** in precisely known aliquots, causing heat to be evolved upon binding. Direct measurement of this heat release with each injection is monitored over time, and determined by the input of power required to maintain equal temperatures between the sample and reference cell. As successive amounts of the KMe3 peptide are titrated into the cell, the quantity of heat released is in direct proportion to the extent of binding. As **$A_2B$** becomes saturated, the heat signal diminishes until only heats of dilution are observed. A binding curve is then obtained by plotting the heats from each injection against the molar ratio of KMe3 and **$A_2B$** in the cell.

New histone 3 peptides were synthesized, containing residues 5-12 from the histone tail as well as an N-terminal tryptophan for concentration determination (Figure 2.24). The tryptophan was separated from the histone sequence by three glycine spacers to prevent any possible interactions between the Trp and either KMe3 or **$A_2B$**. This peptide also contained Gly8 as opposed to Arg8 as found in the native sequence, although this mutation has been shown to have minimal affect on the overall binding affinity. ITC binding curves from the titration of Ac-WGGGQTAG**$KMe_3$**STG-$NH_2$ into *rac*- and *meso*-**$A_2B$** gave dissociation constants of 20 and 13 μM, respectively, confirming that these binding events occur with low micromolar affinity and that the two methods of analysis are comparable.

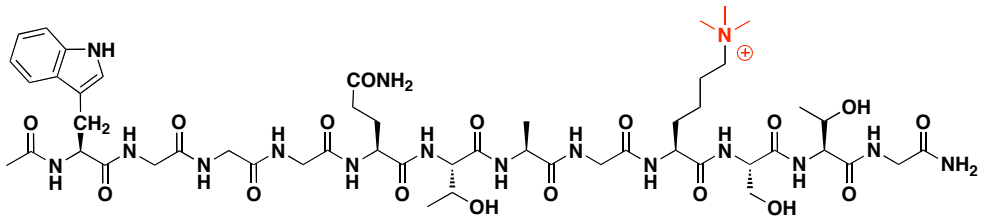


**Figure 2.24.** Histone 3 peptide (Ac-WGGGQTAG**KMe$_3$**STG-NH$_2$) used as the titrant in ITC experiments to confirm **A$_2$B** binding to KMe3.

**i. Structure-function studies.** DCC experiments provide an ideal system for studying the fundamental aspects of molecular recognition, allowing for the systematic variation of both the guest and host molecules. We have shown that varying the methylation state of Lys and Arg guests influences the extent of both amplification and binding, gaining insight into the features required for the selective recognition of these PTMs. It is also feasible to evaluate how variations in the receptor influence binding and selectivity, either positively or negatively, through simple systematic variations of the individual monomers. For example, either the position or spacing of the thiols or the position of the carboxylate relative to the thiols in monomer **B** can be varied to see how these changes influence the amplification of PTM peptides (Figure 2.25).

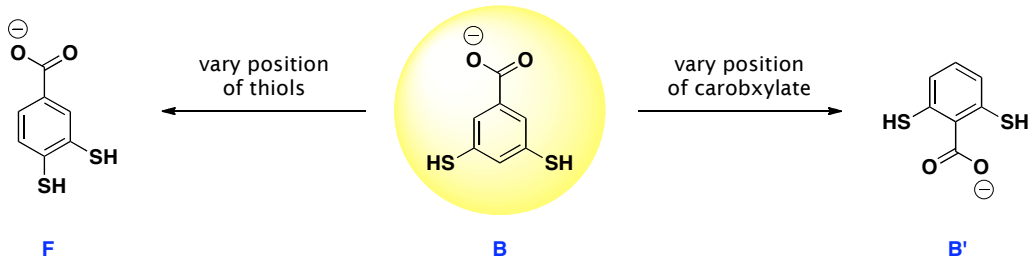
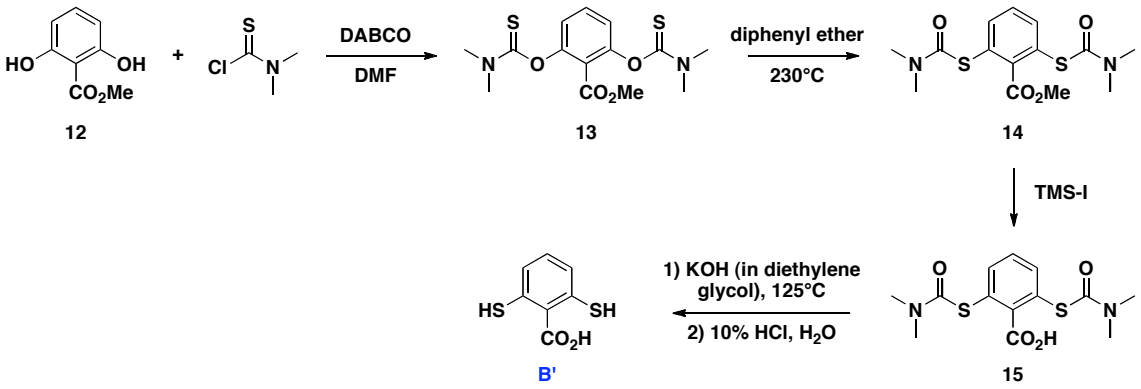



**Figure 2.25.** Possible variations of monomer **B**.

Monomer **B'** was first investigated, where the thiols are placed in the 2- and 6-positions of the ring relative to the carboxylate. Although monomer **B** is somewhat free

to rotate about the disulfide bonds within the **$A_2B$** macrocycle, it is feasible that moving the position of the carboxylate may influence binding. If the carboxylate was previously on the outer face of the receptor and is moved towards the interior of the binding cavity, it is possible that it could facilitate additional favorable electrostatic interactions with positively charged methylated lysine guests. In contrast, there may not be room for the carboxylate in the interior of the cavity, preventing such an interaction, or even getting in the way of binding. By screening monomer **B'** in a biased library against all of the methylated lysine guests, we can gain insight into how this subtle variation influences binding.

Monomer **B'** was synthesized similarly to **B**, however deprotection of the methyl ester was performed first with TMS-I to give **15**,[49] followed by installation of the thiols with KOH (Scheme 2.4). Initial attempts to simultaneously deprotect both protecting groups resulted in incomplete conversion of the methyl ester to the carboxylic acid, so this was circumvented with a two-step deprotection.




**Scheme 2.4.** Synthesis of building block **B'**.

Biased analytical DCLs were setup as described previously by mixing building blocks **A** and **B'** in a 2:1 ratio (7.5 mM total), and the selectivity of the receptors for different

methylation states of lysine and arginine was investigated using dipeptide guests. Upon equilibration, the libraries were analyzed, and the extent of amplification was evaluated by LC/MS. Surprisingly, in the untemplated library only one peak was found corresponding to the mass of **$A_2B'$**, either *rac*-**$A_2B'$** or *meso*-**$A_2B'$**, and it was present in a substantially higher concentration than either of the **$A_2B$** isomers in untemplated DCLs, indicating a higher thermodynamic stability of **$A_2B$**' in comparison to **$A_2B$**. This was a clear indication that monomers **B** and **B'** do not behave identically in the context of these libraries. Overall, the amplification of two different receptors was observed, **$A_2B'$** and **$AB'_3$**, and interestingly, a similar degree of amplification was seen for all lysine methylation states as well as for unmethylated lysine (Figure 2.26). The comparable recognition of Lys and LysMe3 by **$A_2B'$** and **$AB'_3$** indicate that the increased hydrophobicity of LysMe3, its ability to form stronger cation-π interactions, and its dispersion of positive charge over a great surface area are no longer the dominant interactions dictating binding specificity. Instead, the lack of selectivity observed may be the result of additional electrostatic and/or hydrogen bonding interactions with each lysine guest due to the position of the carboxylic acid in **B'** as compared to **B**. In contrast to the lysine PTMs, only minimal amplification of **$A_2B'$** and **$AB'_3$** was observed in the presence of monomethylated and asymmetric dimethylated arginine guests.

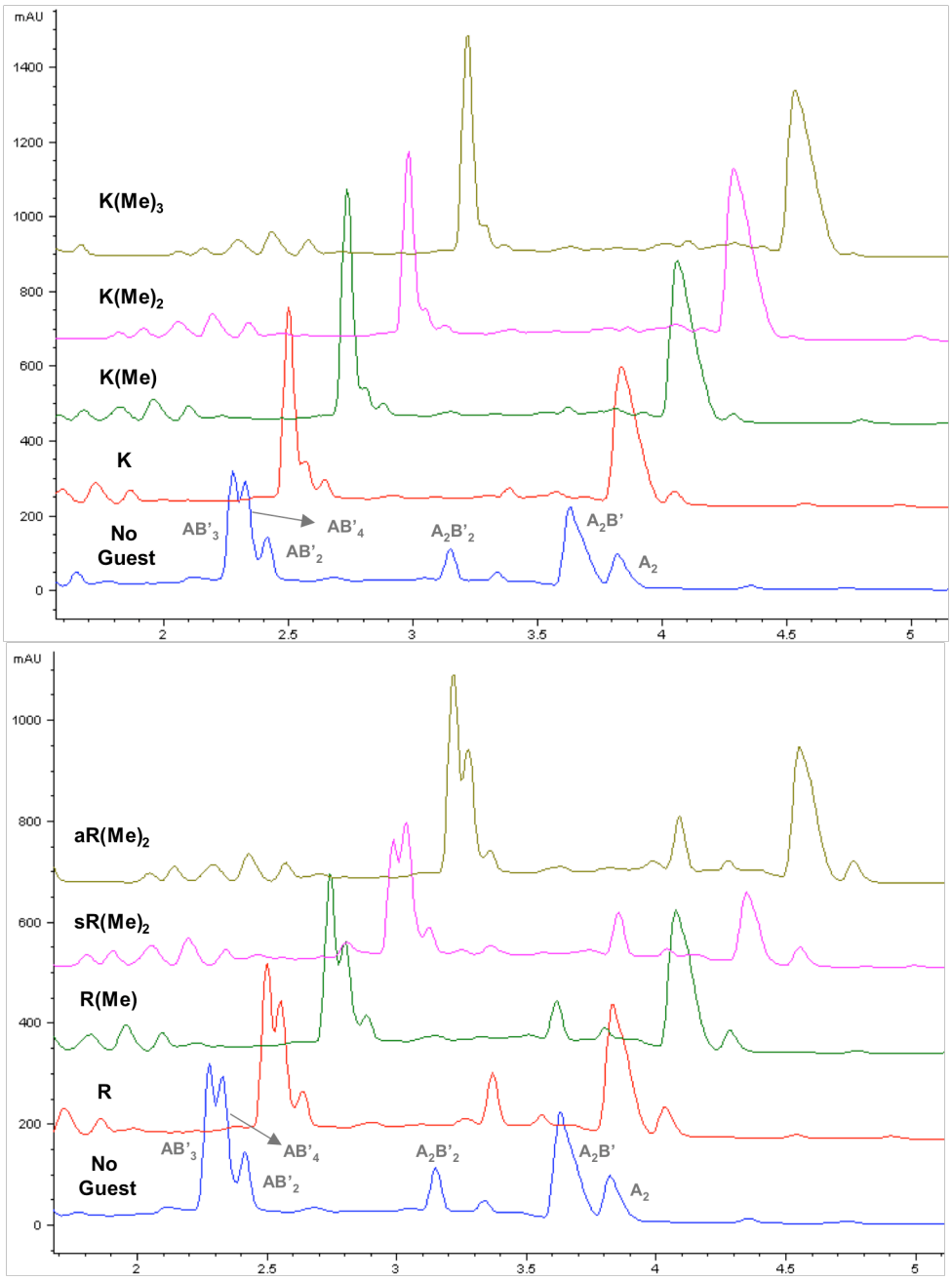



**Figure 2.26.** Part of the analytical HPLC trace at 254 nm of a biased DCL consisting of monomers **A** (5 mM) and **B'** (2.5 mM), untemplated and in the presence of all lysine methylation states (top) and arginine methylation states (bottom).

An **AB'$_3$** biased library was also prepared (7.5 mM monomer total) in attempt to potentially enhance the amplification of this receptor over **A$_2$B'**; however it was discovered that this ratio of monomers formed to a library dominated by macrocycles

containing only **B'**, forming up to a **B'** octamer. Under these conditions only minimal amplification was observed in the presence of KMe3 as seen in Figure 2.27, and this clearly demonstrates the drastic influence that the ratio of monomers has on the library composition and thus on the extent of amplification observed.

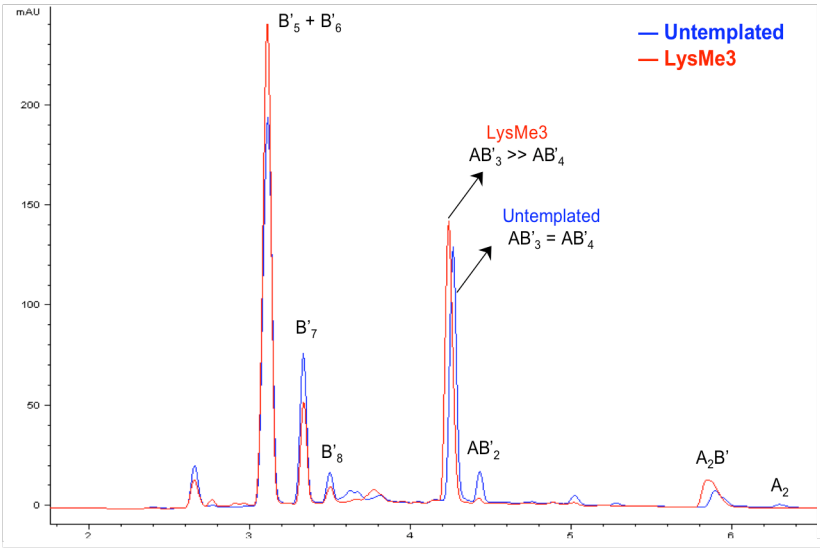



**Figure 2.27.** The analytical HPLC trace at 254 nm of an **AB'$_3$** biased DCL consisting of monomers **A** (1.9 mM) and **B'** (5.9 mM), untemplated (blue) and in the presence LysMe3 (red).

Due to the fact that the equilibrium of all **A$_2$B'** biased libraries with lysine templates was shifted entirely towards two amplified macrocycles, **A$_2$B'** and **AB'$_3$** with only minor traces of other macrocycles, it is difficult to conclude under these conditions that recognition is entirely unselective. To further pursue this observation, libraries were prepared at one tenth the concentration of the original libraries, (0.5 mM **A**, 0.25 mM **B'**, 0.75 mM guest). By lowering the monomer concentrations, we hoped to decrease both the concentration of **A$_2$B'** and **AB'$_3$** in the untemplated library as well as the magnitude of amplification, increasing the overall selectivity. A templated library at lower

concentrations is also more likely to selectively amplify the better host, which would potentially provide insight as to whether **$A_2B'$** or **$AB'_3$** is the better receptor of the two.

Analysis of these lower concentration libraries revealed only modest **$A_2B'$** selectivity for increasing lysine methylation states, with about equal amplification of KMe2 and KMe3 (Figure 2.28). As expected, minimal **$A_2B'$** amplification was observed for any of the methylated arginine templates. Both **$A_2B'$** and **$AB'_3$** receptors remained dominant species in the untemplated library resulting in a fairly small percent change in concentration upon amplification. Amplification of **$AB'_3$** was not accurately assessed in this case, as poor chromatographic resolution did not allow for separation of **$AB'_3$** from **$AB'_4$**. Although it was apparent from the mass spectra of this peak that the ratio of these two macrocycles differed in the untemplated and templated libraries, accurate integrations were not feasible. Unbiased libraries with a 1:1 ratio of **A**:**B'** (0.25 mM each) were also prepared and analyzed in attempt to achieve further selectivity, however these libraries revealed amplification almost identical to that of the biased library. Although the absolute binding affinity of these new receptors has not been pursued further, as macrocycles which also bind unmethylated lysine are of much less value, this simple structure function study through the modification of one functional group on a single monomer provided interesting results and further insight as to the factors contributing to KMe3 recognition. The difference in selectivity observed when **B** is replaced with **B'** is surprising, and demonstrates that subtle structural changes can have a significant effect on binding.

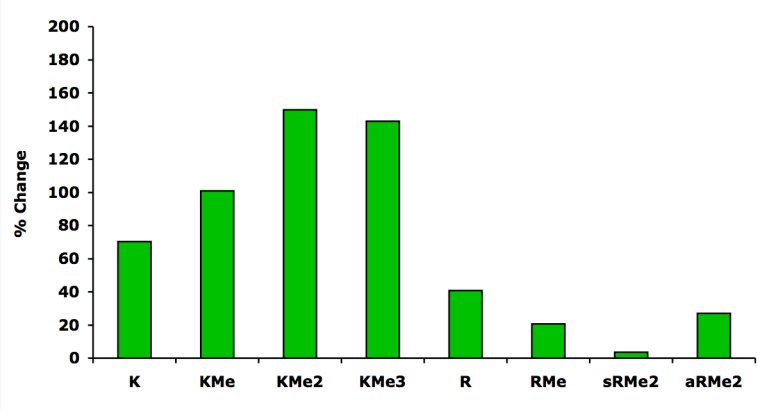



**Figure 2.28.** Amplification results of the lower concentration **$A_2B$'** library (0.75 mM total) templated with all lysine and arginine methylation states, represented as a percent change in peak area.

In addition to changing the position of the carboxylate, the position of the thiols can also be varied relative to each other and relative to the carboxylate. We chose to investigate monomer **F** (Figure 2.25), where the thiols are placed in closer proximity on adjacent carbons. A library was prepared with a 1:1:1 ratio of monomers **A**, **F**, and **H** (7.5 mM total), with the non-aromatic dithiol monomer **H** (Figure 2.12) included largely to serve as a linker to help cyclize molecules which would otherwise be too strained to cyclize. In contrast to the change in selectivity observed when **B** is replaced with **B'**, when **B** is replaced with monomer **F** the amplification trends seem to mimic that of **$A_2B$**. Two **$A_2F$** isomers are amplified significantly in the presence of KMe3, and both isomers exist in very low concentrations in the untemplated library (Figure 2.29). The amplification of the **$A_2F$** macrocycle at 8.5 minutes is also greater than it appears, as the receptor co-elutes with other macrocycles in the untemplated library only, increasing the area of this peak. Upon templation, these other co-eluting macrocycles are depleted. This study clearly shows that while some monomer structural variations result in a fairly drastic change in the extent of amplification and selectivity between PTMs, others have

little impact on binding. Many other structural variations of monomer **B** can be envisioned, and there are ongoing investigations in the Waters Lab to pursue such structure-function studies further.

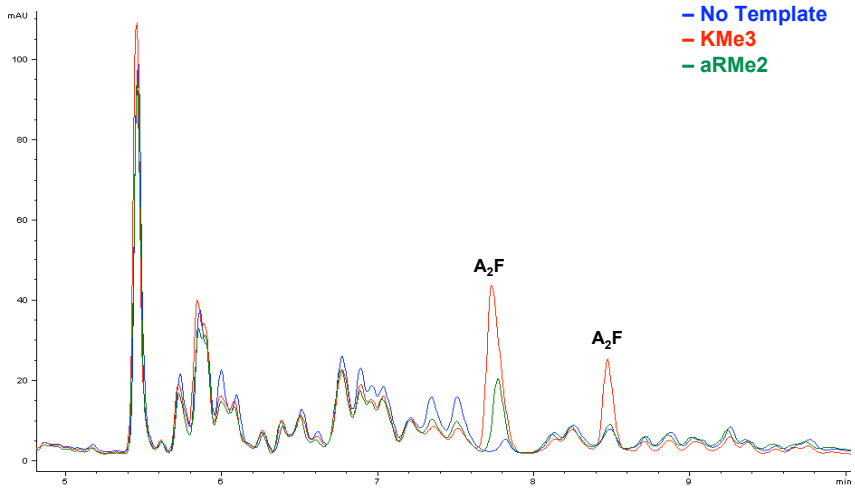



**Figure 2.29.** The analytical HPLC trace at 254 nm of a DCL consisting of monomers **A** (2.5 mM), **F** (2.5 mM), and **H** (2.5 mM) untemplated (blue) and in the presence KMe3 (red) and aRMe2 (green).

### ii. Conclusions

In conclusion, we have identified synthetic receptors for KMe3 that exhibit both comparable affinity and selectivity to the native HP1 chromodomain. The mass of $\mathbf{A_2B}$ is less than 900 Da, as compared to approximately 6300 Da for the HP1 chromodomain and about 150 kDa for a typical antibody. The comparable binding affinity to the native protein is impressive, particularly given that the synthetic receptor appears to bind only to the KMe3 sidechain, whereas the chromodomain binds to the surrounding sequence as well. The observed selectivity for higher methylation states with $\mathbf{A_2B}$ can be attributed to differences in the magnitude of the cation-π interactions, as well as differences in size and desolvation penalty. The difference in selectivity observed when **B** is replaced with **B'** is surprising, and demonstrates that subtle structural changes can have a significant

effect on binding. Such sequence-independent receptors have the potential to be widely used in the investigation of unknown protein Lys methylation sites,[50] unlike antibodies which are sequence selective. This work suggests that small molecule receptors identified via DCC have a promising future as affinity reagents for PTMs, as they can easily be synthetically modified with a fluorophore, cell-penetrating peptide, or other tag.

### iii. Experimental Section

**a. Monomer Synthesis.** Monomers **A**, **B**, and **F** were synthesized following literature procedures.[43,51] Monomer **H** is commercially available. Monomer **C** was synthesized following a modified procedure of Field and Kirrstetter.[44b]

*Methyl 3,5-dimethylbenzoate (**10**)*. To a solution of 3,5-dimethylbenzoic acid **9** (5.0 g, 33.3 mmol) in 50 mL of methanol, sulfuric acid (10 ml, 184 mmol) was slowly added, producing a significant amount of heat. The solution was then refluxed overnight under nitrogen at about 85°C. The reaction was monitored by TLC (silica gel, 5% MeOH in $CH_2Cl_2$ with TFA; $R_f$ starting material: 0.5; $R_f$ product: 0.83). The colorless reaction mixture was allowed to cool and the methanol was removed by rotary evaporation. The remaining solution was then poured into a 300 ml beaker of ice water to obtain a white precipitate which extracted into ether, washed with $NaHCO_3$, dried with $MgSO_4$. The $MgSO_4$ was removed by filtration and the solvent was evaporated to give a light yellow oil. The product was recrystallized by adding petroleum ether (2-3 ml) to give 5.1 g of a clear, colorless solid (31.2 mmol, 94%). $^1$H NMR ($CDCl_3$, 400 MHz): $\delta$ = 7.63 (s, 2H), 7.18 (s, 1H), 3.90 (s, 3H), 2.34 (s, 6H).

*Methyl 3,5-bis(bromomethyl)benzoate (**11**)*. To a solution of methyl 3,5-dimethylbenzoate **10** (4.4 g, 27.1 mmol) in 60 ml of carbon tetrachloride *N*-bromosuccinimde (9.7 g, 54.2 mmol) was added. Peroxide (0.13 g, 0.54 mmol) was then added to the suspension, and it was refluxed for one hour. The reaction was monitored by TLC (silica gel, 10% EtOAc in hexanes; $R_f$ starting material: 0.70; $R_f$ product: 0.35). The reaction was cooled and the NBS was removed by filtration. The solution was then washed with water, dried with $MgSO_4$, and the $MgSO_4$ was removed by filtration. The solvent was removed by evaporation, and the product was precipitated overnight in petroleum ether in the fridge. If necessary the crude product can be recrystallized in ether to give pure white crystals (12.2 mmol, 45%). $^1$H NMR ($CDCl_3$, 400 MHz): δ = 7.99 (d, 2H), 7.60 (t, 1H), 4.48 (s, 4H), 3.92 (s, 3H).

*3,5-bis(mercaptomethyl)benzoic acid (**C**)*. A solution of methyl 3,5-bis(bromomethyl)benzoate **11** (0.74 g, 2.3 mmol) and thiourea ( 0.46 g, 6.0 mmol) in 10 ml of 95% ethanol was refluxed for 6 hours under nitrogen, after which no starting material remained as observed by TLC (10% EtOAc in hexanes). To the reaction 9.7 mL of 1M NaOH was added, and the solution was refluxed overnight. The reaction was cooled and acidified with 10% HCl while stirring to give a cloudy white solution. The solution was filtered to give a white, fluffy solid (1.8 mmol, 78%). $^1$H NMR ($(CD_3)_2CO$), 400 MHz): δ = 7.90 (d, 2H), 7.61 (t, 1H), 3.84 (d, 4H), 2.31 (t, 2H, SH).

*2,6-Bis(dimethylthiocarbamoyloxy)benzoic acid methyl ester (**13**)*. Methyl 2,6-dihydroxybenzoate **12** (2.0 g, 11.9 mmol) was dissolved in anhydrous DMF (8 mL) under a nitrogen atmosphere. The solution was cooled to 0 °C and DABCO (8.0 g, 71.4) was added in portions. To the resulting suspension a solution of *N,N*-dimethylthiocarbamoyl

chloride (8.9 g, 71.4 mmol) in DMF (8 mL) was added dropwise. In cases where the solution was too thick to stir, additional DMF was added. The suspension was allowed to warm to room temperature overnight. The reaction mixture was poured into water (80 mL) and stirred before filtering. The product was washed with water and then ethanol to give a pure, white solid (2.8 g, 68%). $^1$H NMR ($CDCl_3$, 400 MHz): δ = 7.49 (t, 1H), 7.03 (d, 2H), 3.74 (s, 3H), 3.42 (s, 3H), 3.33 (s, 3H).

*2,6-Bis(dimethylcarbamoylsulfanyl)benzoic acid methyl ester (**14**)*. O-thiocarbamate **13** (0.5 g, 1.46 mmol) was suspended in diphenyl either (5 mL) and heated in a sand bath to 230-240 °C under a nitrogen atmosphere. Once the appropriate temperature was reached, the reaction was stirred for an additional 3 hours. The reaction was monitored by TLC (EtOAc, $R_f$ starting material: 0.58, $R_f$ product: 0.33). The reaction mixture was allowed to cool and then purified by column chromatography (100% hexane to elute phenyl ether, then 1:1 EtOAc;hexane followed by 7:3 EtOAc:hexane). The solvent was evaporated and the compound was dried to give a beige solid 0.35 g, 70%). $^1$H NMR ($CDCl_3$, 400 MHz): δ = 7.60 (d, 2H), 7.45 (t, 1H), 3.88 (s, 3H), 3.02 (broad d, 6H).

*2,6-Bis(dimethylcarbamoylsulfanyl)benzoic acid (**15**)*.[49] A mixture of 2,6-bis(dimethylcarbamoylsulfanyl)benzoic acid methyl ester **14** (0.16 g, 0.47 mmol) and $Me_3SiI$ (0.36 mL, 2.52 mmol) were refluxed at 105°C overnight. Once cool, diethyl ether (11 mL) was added and washed with 10% NaOH. The basic aqueous layers were combined and 1M HCl was added to acidify the solution. A white crystal-like solid precipitated out and was filtered (0.084 g, 0.26 mmol). $^1$H NMR (MeOD, 400 MHz): δ = 7.59 (d, 2H), 7.42 (t, 1H), 3.07 (broad d, 12H).

*2,6-Dimercaptobenzoic acid (**B'**).* Under a nitrogen atmosphere 2,6-bis(dimethylcarbamoylsulfanyl)benzoic acid **15** (0.037 g, 0.011 mmol) was suspended in a 1.75 M solution (0.38 mL) of KOH in diethylene glycol that had been purged with argon for 2 hr. The solution was heated at 105°C for 30 min. After the solution had cooled to room temperature, 2.7 mL of purged water was added followed by rapid addition of 10% HCl (0.5 mL). The product was either filtered when possible and washed with hexanes, or diluted further and purified by reverse phase HPLC (0.13 g, 0.070 mmol). $^1$H NMR (MeOD, 400 MHz): $\delta$ = 7.19 (d, 2H), 7.07 (t, 1H).

**b. Synthesis of methylated lysine and arginine peptides and fluorescent labeling.** All peptide synthesis was performed on a Tetras Peptide Synthesizer using Peptides International CLEAR-Amide resin. Peptides were synthesized on a 0.06 mmol scale by standard Fmoc solid phase synthesis. All amino acids with functionality were protected during synthesis. Coupling reagents were HOBt/HBTU in DMF. For the dipeptides and the histone tail peptides used for templation studies, as well as the histone peptides synthesized for ITC experiments, the N-terminus was acylated with a solution of 5% acetic anhydride and 6% 2,6-lutidine in DMF. Peptides synthesized for fluorescence anisotropy were capped with 2 equivalents of 5(6)-carboxyfluorescein and coupled in the dark with standard coupling reagents overnight. Cleavage was performed by hand with a cocktail of 95% TFA/2.5% triisopropylsilane/2.5% $H_2O$ for 3 hours. Peptides were purified by semipreparative reversed-phase HPLC on a C18 column at a flow rate of 4 mL/min. Peptides were purified with a linear gradient of A and B (A: 95% $H_2O$/5% $CH_3CN$ with 0.1% TFA, B: 95% $CH_3CN$/5% $H_2O$ with 0.1% TFA) and elution was

monitored at 214 nm. Once purified, peptides were lyophilized to powder and characterized by ESI-MS.

All PTM peptides were synthesized using monomethylated (Fmoc-Lys(Me)(Boc)-OH, Fmoc-Arg(Me)(Pbf)-OH) and dimethylated (Fmoc-LysMe2-OH, symmetric or asymmetric Fmoc-ArgMe2-OH) lysine and arginine residues which are commercially available from Anaspec, Bachem, or Novabiochem. The coupling of methylated amino acids was done with 2 equivalents of the unnatural amino acid for 10-12 hours. The trimethyllysine-containing peptides were synthesized by reacting the corresponding dimethylated peptides (0.6 mmol scale) prior to cleavage from the resin with MTBD (10.8 μL, 0.075 mmol) and methyl iodide (37.4 μL, 0.6 mmol) in DMF (5 mL) for 5 hours with bubbling $N_2$ in a peptide synthesis flask stoppered with a vented septum. After washing the resin with DMF (3x), $CH_2Cl_2$ (3x), and drying, the peptides were cleaved, purified, and characterized as normal.

**c. Dynamic combinatorial library screens.** The relevant building blocks were individually dissolved in water, adding sufficient 1.0 M aqueous NaOH to fully deprotonate the thiols and carboxylic acids on the building blocks, using sonication when necessary. The pH of each solution was then adjusted to 8.5 using 1.0 M aqueous HCl and 1.0 M aqueous NaOH. For example, in the unbiased **ABC** libraries, aliquots of each monomer solution were combined in a 2 mL LC-MS vial to reach a final concentration of 2.5 mM of each monomer. In the biased $\mathbf{A_2B}$ and $\mathbf{A_2B'}$ libraries, aliquots of each monomer solution were combined in a 2 mL LC-MS vial to reach a final concentration of 5 mM **A** and 2.5 mM of **B** or **B'** (unless noted otherwise). When necessary, aliquots of peptide guests dissolved in water were added to the reactions to reach a final

concentration of 7.5 mM peptide. The peptide guests were not found to significantly alter the reaction pH, however if necessary, the pH was adjusted again after addition of the guests. Any remaining volume was made up with water. The vials were capped and analyzed at various time points.

**d. Analytical LC/MS.** LC-MS was carried out on an Agilent Rapid Resolution LC-MSD system equipped with an online degasser, binary pump, autosampler, heated column compartment, and diode array detector. All separations were performed using 5 mM $NH_4OAc$ $H_2O$-acetonitrile gradients at pH 5 and a Halo C18 column (4.6 × 100 mm, 2.7 micron). The MS was performed using a single quad mass spectrometer. Mass spectra (ESI-) were acquired in ultrascan mode by using a drying temperature of 350°C, a nebulizer pressure of 45 psi, a drying gas flow of 10 L/min, and a capillary voltage of 3000 V. The reactions were monitored weekly (3 μL injections) until equilibrium was reached after about 3 weeks. Larger injection volumes (10 μL) were used for the lower concentration libraries (0.75 mM total monomer) for adequate MS detection. The chromatography of the **ABC** DCL was carried out at 50°C with gradient A (Table 2.3). The chromatography of library $\mathbf{A_2B}$ was carried out with gradient B using a gradient temperature, going from 50°C to 40°C, left to right. The chromatography of library $\mathbf{A_2B'}$ was carried out at 50°C with gradient C. The peak areas were integrated at 254 nm and the amplification factors were calculated (A.F. = % area of $\mathbf{A_2B}$ in templated DCL / % area of $\mathbf{A_2B}$ in untemplated DCL).

**Table 2.3.** Analytical LC methods use to analyze DCC libraries **ABC** (Method A), **$A_2B$** (Method B) and **$A_2B'$** (Method C).

| **Gradient A** | | |
|---|---|---|
| **Time (min)** | **%B** | **Flow Rate (mL/min)** |
| 0.00 | 0.0 | 1.0 |
| 3.00 | 30.0 | 1.0 |
| 7.00 | 32.3 | 1.0 |
| 7.30 | 32.3 | 1.0 |
| 7.35 | 32.3 | 0.6 |
| 9.00 | 32.3 | 0.6 |
| 9.10 | 32.3 | 1.0 |
| 10.80 | 34.0 | 1.0 |
| 10.90 | 100.0 | 1.0 |
| 11.90 | 100.0 | 1.0 |
| 12.00 | 0.0 | 1.0 |
| 13.00 | 0.0 | 1.0 |

| **Gradient B** | | |
|---|---|---|
| **Time (min)** | **%B** | **Flow Rate (mL/min)** |
| 0.00 | 3.0 | 1.0 |
| 3.00 | 30.0 | 1.0 |
| 8.25 | 33.0 | 1.0 |
| 8.30 | 33.0 | 1.5 |
| 10.30 | 34.0 | 1.5 |
| 12.00 | 50.0 | 1.5 |
| 12.10 | 100.0 | 1.5 |
| 13.50 | 100.0 | 1.5 |
| 14.00 | 3.0 | 1.0 |
| 19.00 | 3.0 | 1.0 |

| **Gradient C** | | |
|---|---|---|
| **Time (min)** | **%B** | **Flow Rate (mL/min)** |
| 0.00 | 15.0 | 1.5 |
| 2.00 | 30.0 | 1.5 |
| 6.00 | 30.0 | 1.5 |
| 9.00 | 50.0 | 1.5 |
| 9.10 | 100.0 | 1.5 |
| 11.00 | 100.0 | 1.5 |
| 11.50 | 15.0 | 1.5 |
| 16.00 | 15.0 | 1.5 |

**e. Synthesis and isolation of *rac*-$A_2B$ and *meso*-$A_2B$.** Biased libraries were prepared on a 20 mL scale (**A**: 35.6 mg, 0.1 mmol; **B**: 9.3 mg, 0.05 mmol) templated with methylisoquinoline triflate (44.0 mg, 0.15 mmol). Upon equilibration the libraries were neutralized and the receptors were isolated by semi-preparative HPLC. Approximately 0.5 mL injections were chromatographed using standard peptide synthesis mobile phases A and B (0-50% B 0-5 min, then isocratic at 50% B 5-20 min) with a flow rate of 4.0 mL/min. Optimal separation of the two isomers was achieved using a column heater set to 40°C. The two $\mathbf{A_2B}$ peaks from 11.5 – 12.5 minutes were collected separately (Figure 2.30) and analyzed for purity by analytical LC-MS (Figure 2.31), as their retention times differ by about 0.7 minutes. Both peaks are indistinguishable by mass (Figure 2.32). Purified *rac*- and *meso*-$\mathbf{A_2B}$ were lyophilized to powder and stored under nitrogen, however in some cases a second purification was required to achieve pure *meso*-$\mathbf{A_2B}$.

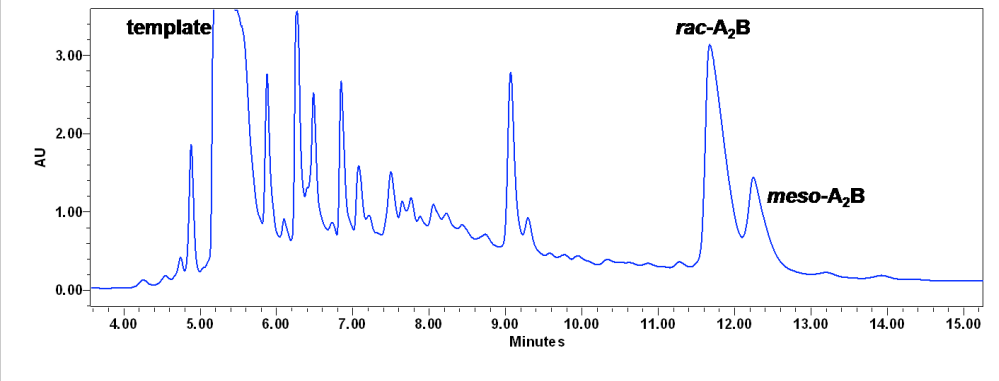



**Figure 2.30.** Semi-preparative HPLC trace of an $\mathbf{A_2B}$ biased library.

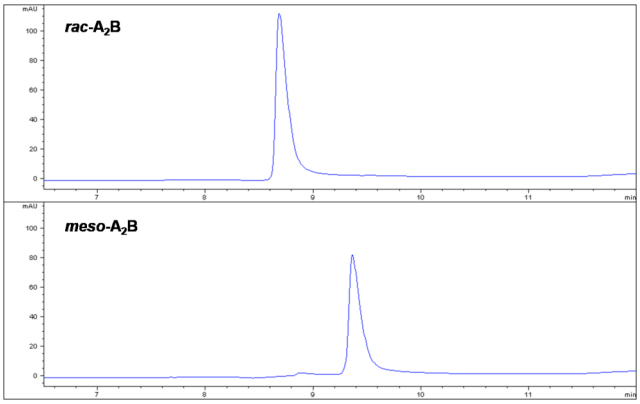



**Figure 2.31.** Analytical LC traces of purified *rac*-**A$_2$B** (top) and *meso*-**A$_2$B** (bottom).

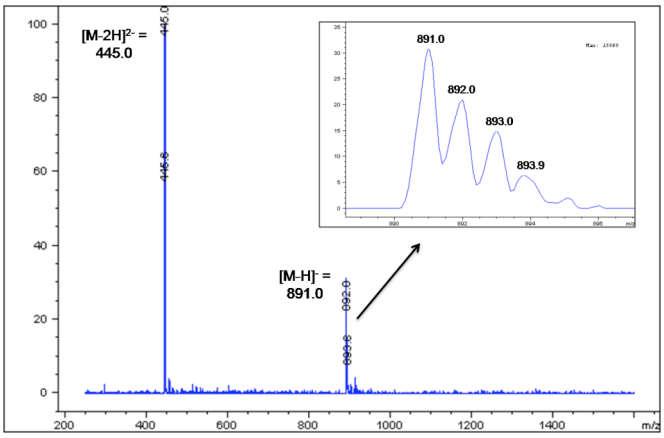



**Figure 2.32.** Mass spectra of both purified *rac*-**A$_2$B** and *meso*-**A$_2$B** (-ESI).

**f. NMR spectroscopy.** All NMR samples were dissolved in 50 mM borate buffer at approximately pD 9 containing 0.05 mM DSS. To prepare the NMR buffer $B_4Na_2O_7$ was dissolved in $D_2O$. The buffer was either adjusted to pD 9 with acetic acid-$d_4$, or used without any further adjustment, giving a pD of 9.25. An NMR sample of the H-K$Me_3$-G-$NH_2$ dipeptide was prepared with a concentration of approximately 1 mM. A second sample was prepared containing both *rac*-**A$_2$B** (0.85 mM) and KMe3 (0.65 mM), with the *rac*-**A$_2$B** in excess to ensure all of the trimethyllysine was bound. Samples were analyzed

on a Varian Inova 600-MHz instrument at either 25 °C or 5 °C. 1D NMR spectra were collected using 32K data points and 128 scans using a 1-3 second presaturation. 2D TOSCY experiments were taken with 16-64 scans in the first dimension and 128-256 scans in the second dimension. All spectra were analyzed using standard window functions (sinbell and Gaussian with shifting). Presaturation was used to suppress the water resonance. Lysine assignments were made using standard methods as described by Wurthrich.[52]

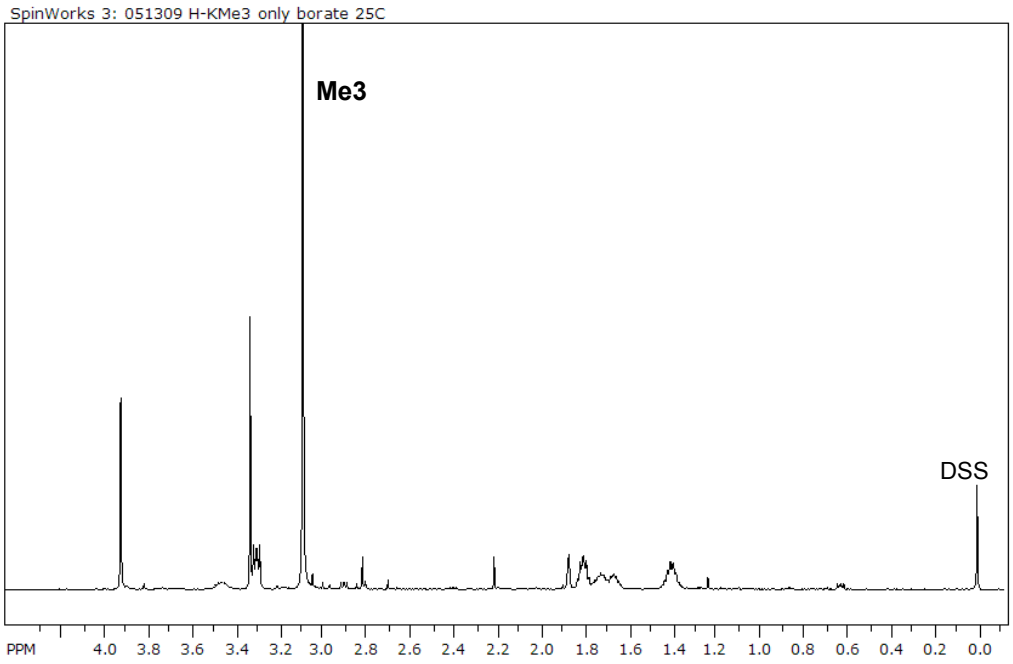



**Figure 2.33.** $^{1}$H NMR of H-KMe$_3$-G-NH$_2$ at 25 °C.

**Table 2.4.** Proton chemical shift assignments for $^{1}$H NMR of H-KMe$_3$-G-NH$_2$.

| Residue | α | β | γ | δ | ε | Me |
|---|---|---|---|---|---|---|
| **KMe$_3$** | 3.47 | 1.73, 1.70 | 1.40 | 1.81 | 3.31 | 3.10 |
| **G** | 3.93 | | | | | |

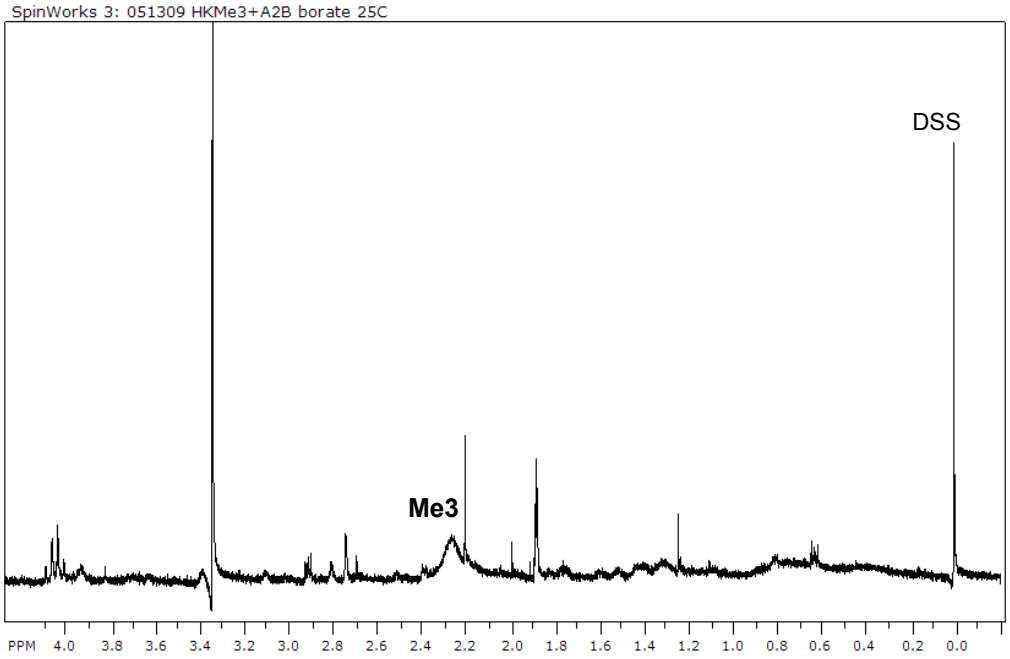



**Figure 2.34.** $^{1}$H NMR of H-K$Me_3$-G-$NH_2$ bound to *rac*-**$A_2$B** at 25 °C. The lysine methyl groups are broadened and shift upfield to 2.27 ppm. The signals at 2.21, 1.89, and 1.25 are due to additional DSS protons and the signal at 3.34 is residual methanol.

**g. $A_2$B extinction coefficient determination.** For all studies investigating the affinity of **$A_2$B** (NMR, anisotropy, ITC), the concentration of **$A_2$B** was determined by either mass or UV-Vis. For the cases in which UV-Vis was used, an extinction coefficient was first determined for the receptor. Due to limited amounts of material of **$A_2$B**, an extinction coefficient was determined for both **A** and **B** individually and used in an additive manner to determine that of the receptor. Stock solutions of a known concentration (between 5 and 10 mM) were prepared for each monomer in aqueous solution titrated to pH 8.5 with 1 M NaOH (with sonication if needed to help dissolve solid). The stock solutions sat for 1-3 weeks to form various single building block macrocycles, eliminating most free thiolate anions, and therefore more closely resembling the UV-Vis spectra of the monomers within the **$A_2$B** macrocycle.

Various successive dilutions of each monomer were made and the UV-Vis spectra was taken at each concentration four times and averaged. The spectra were taken on a Thermo Scientific NanoDrop Spectrophotometer, using 2 μl for each measurement. The absorbance at 315 nm was plotted against concentration with the slope set to zero, and based on Beer's Law, the slope was calculated to determine the extinction coefficient of each monomer. This was repeated 2 to 3 times for each monomer, with some slight variance between experiments, and the slopes giving the highest $R^2$ values were used to determine the extinction coefficients. For monomer B the extinction coefficient was determined to be 1463 $M^{-1}cm^{-1}$ based on the average of two separate runs giving slopes of 1477 $M^{-1}cm^{-1}$ and 1448 $M^{-1}cm^{-1}$ (Figure 2.35). For monomer A (racemic mixture), the extinction coefficient was determined to 1441 $M^{-1}cm^{-1}$ (Figure 2.36). Therefore, the extinction coefficient for **$A_2B$**, assuming the absorbance of both *rac* and *meso* isomers is indistinguishable, was determined to be 4345 $M^{-1}cm^{-1}$.

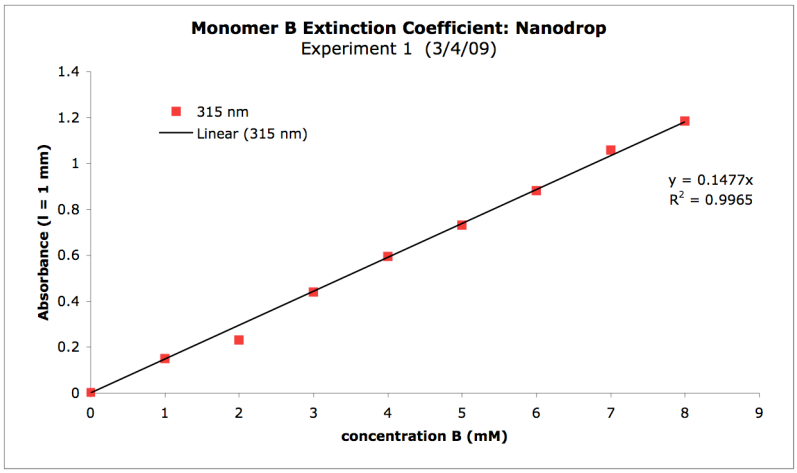



**Figure 2.35.** Determination of the extinction coefficient of monomer **B**.

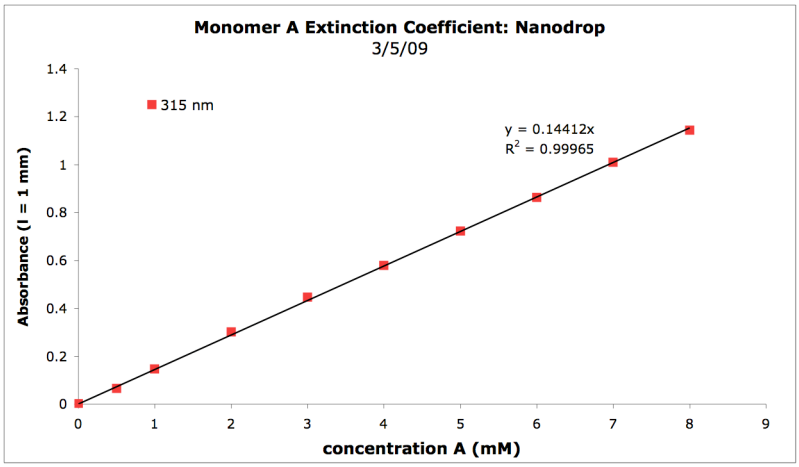



**Figure 2.36.** Determination of the extinction coefficient of monomer **A**.

**h. Fluorescence anisotropy binding experiments.** Binding assays were performed with purified *rac*- or *meso*-$\mathbf{A_2B}$ and fluorescein labeled histone 3 peptides. Peptides were dissolved in buffer (10 mM phosphate pH 8.5), and 100 μM solutions were prepared as determined by UV-Vis at 492 nm ($\varepsilon$ = 78,000 $M^{-1}cm^{-1}$). Stock solutions of $\mathbf{A_2B}$ were prepared, with the solution concentrations determined by either mass or UV-Vis at 315 nm ($\varepsilon$ = 4345 $M^{-1}cm^{-1}$) on a NanoDrop Spectrophotometer. Assays were prepared in 384-well plates (Corning) with a total volume of 50 μL per well, containing 20 μM labeled peptide and increasing concentrations of $\mathbf{A_2B}$ in buffer. Plates were spun down and allowed to incubate for at least 30 minutes before analysis. Fluorescence anisotropy was measured on a PHERAstar plate reader (BMG Labtech) using FP485, 520A, and 520B filters at approximately 27°C. The anisotropy data was plotted as a function $\mathbf{A_2B}$ concentration and each plot was fitted in KaleidaGraph to the following equation:[53]

$$r = \left( \left( \frac{(a + x + k_d) \pm \sqrt{(-a - x - k_d)^2 - 4(a \cdot x)}}{2 \cdot a} \right) \cdot (r_\infty - r_o) \right) + r_o$$

where r is fluorescence anisotropy, $r_o$ is the initial anisotropy value, $r_\infty$ is the maximum anisotropy value, a is the peptide concentration, x is the concentration of **$A_2B$**, and $k_d$ is the dissociation constant. All measurements were taken in duplicate or triplicate.

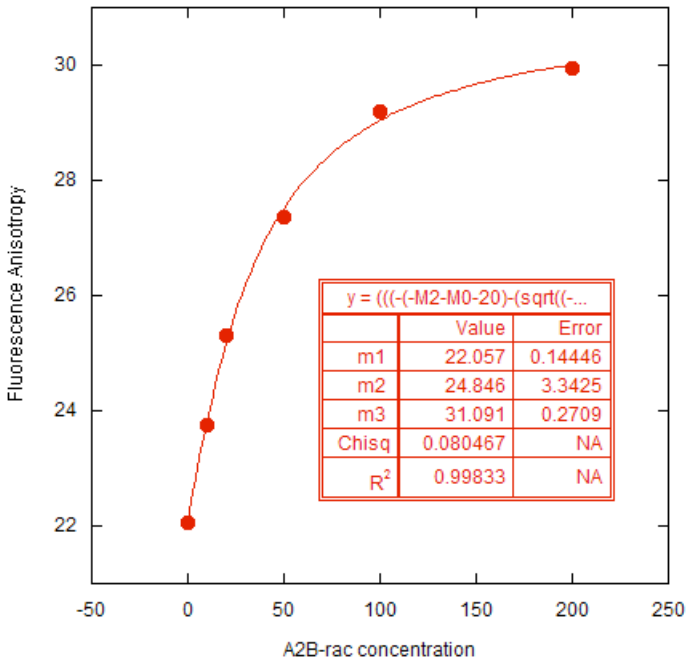



**Figure 2.37.** Fluorescence anisotropy of *rac*-**$A_2B$** with H3 $KMe_3$.

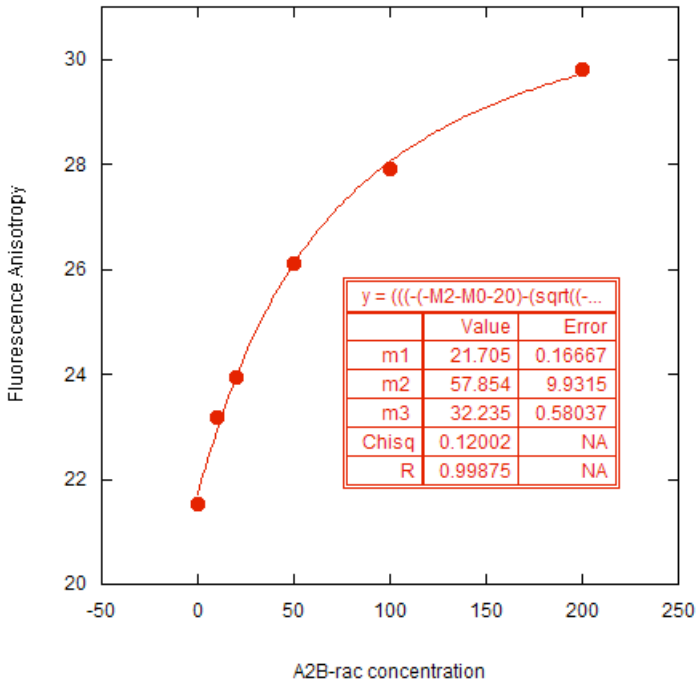



**Figure 2.38.** Fluorescence anisotropy of *rac*-$\mathbf{A_2B}$ with H3 $KMe_2$.

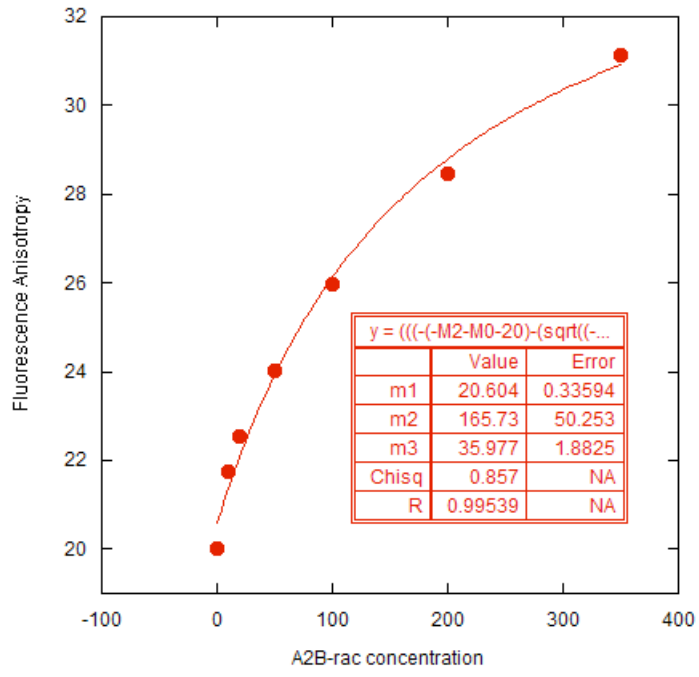



**Figure 2.39.** Fluorescence anisotropy of *rac*-$\mathbf{A_2B}$ with H3 KMe.

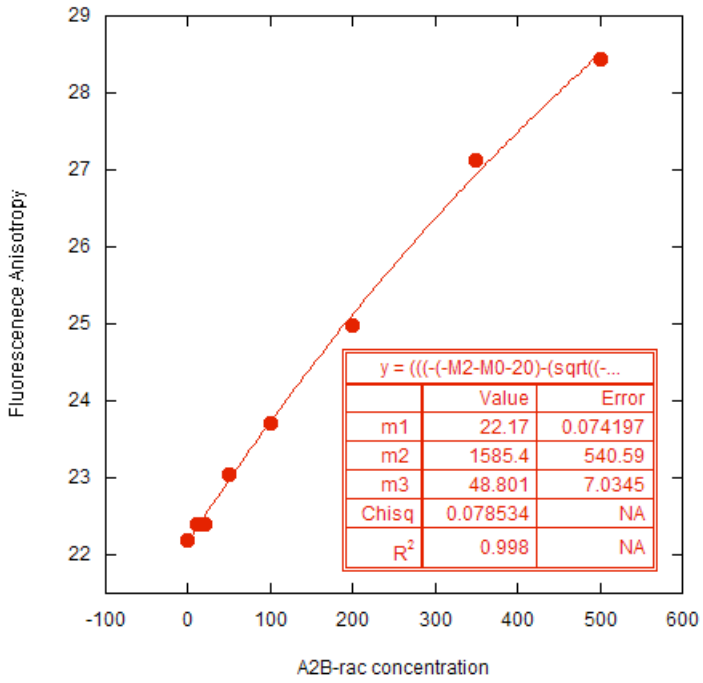



**Figure 2.40.** Fluorescence anisotropy of *rac*-$\mathbf{A_2B}$ with H3 K.

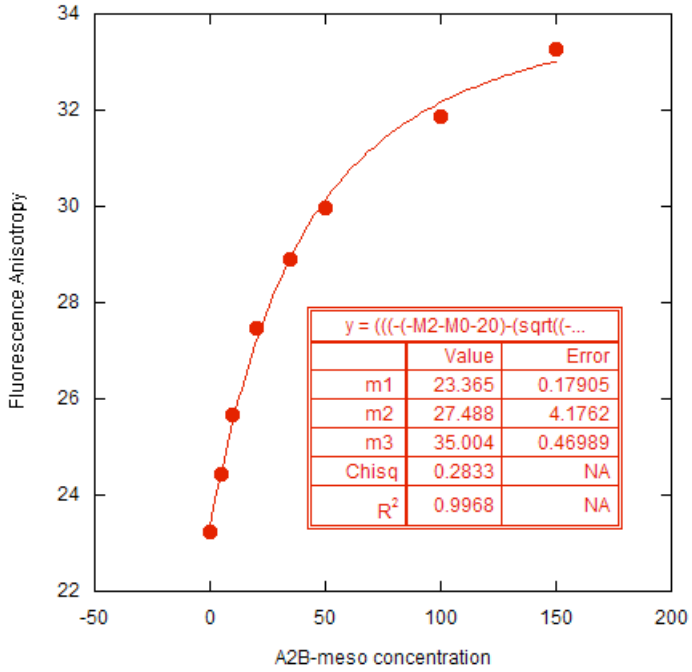



**Figure 2.41.** Fluorescence anisotropy of *meso*-$\mathbf{A_2B}$ with H3 $KMe_3$.

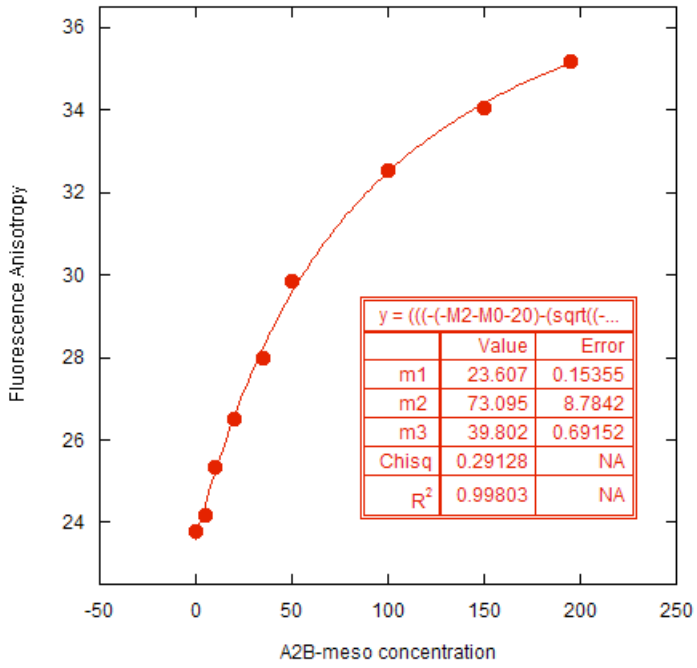



**Figure 2.42.** Fluorescence anisotropy of *meso*-$\mathbf{A_2B}$ with H3 $KMe_2$.

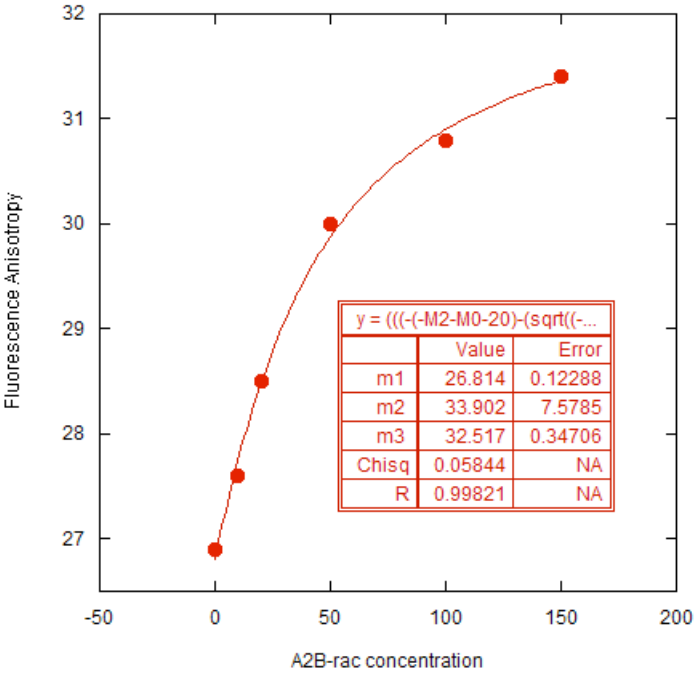



**Figure 2.43.** Fluorescence anisotropy of *rac*-$\mathbf{A_2B}$ with H3 $KMe_3$ where Arg8 has been mutated to Gly8.

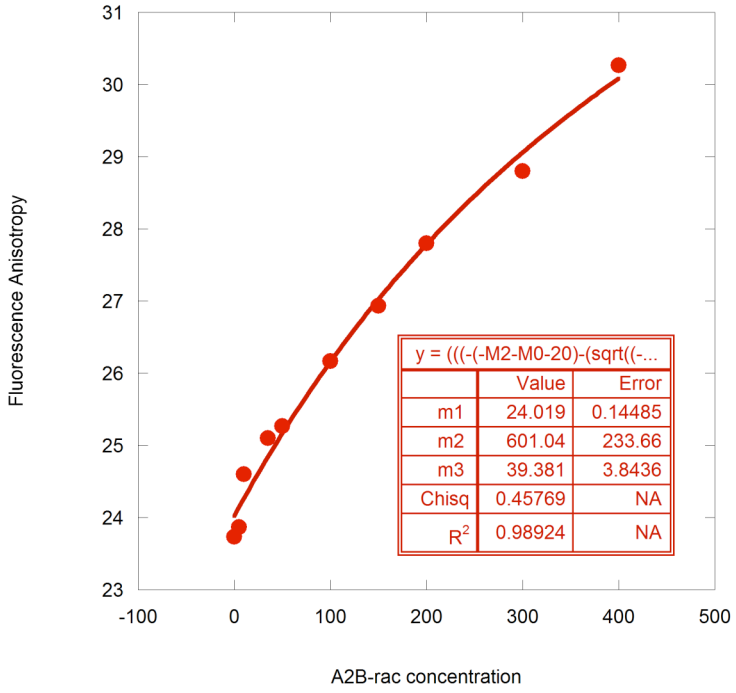



**Figure 2.44.** Fluorescence anisotropy of *rac*-$\mathbf{A_2B}$ with H3 aRMe2.

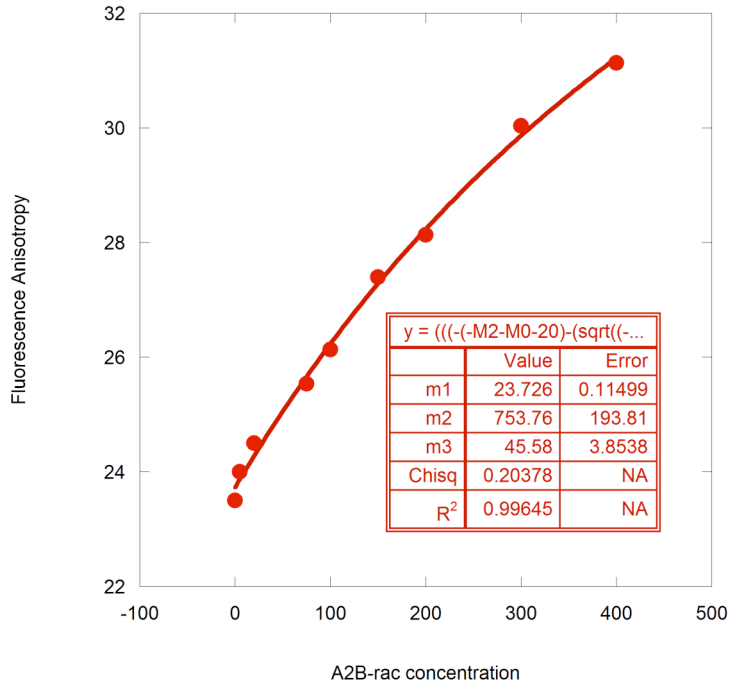



**Figure 2.45.** Fluorescence anisotropy of *rac*-$\mathbf{A_2B}$ with H3 sRMe2.

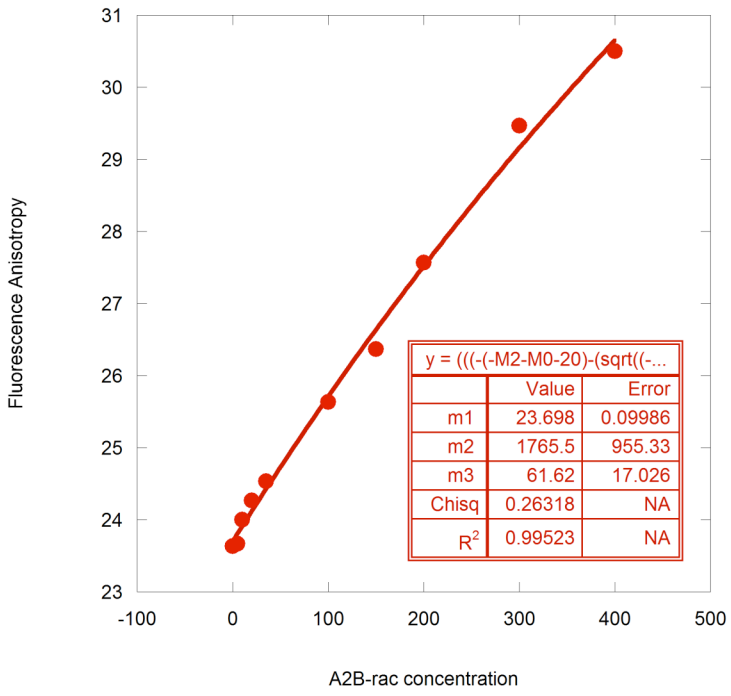



**Figure 2.46.** Fluorescence anisotropy of *rac*-**$A_2B$** with H3 RMe.

**i. Isothermal titration calorimetry binding experiments.** Isothermal titration calorimetry measurements were conducted using an ITC200 from MicroCal, LLC to verify the binding data obtained by fluorescence anisotropy. Solutions were degassed for 10 minutes under vacuum prior to the experiment. The titrant used to determine the binding to $KMe_3$ was a 2.0 mM solution of the H3 tail peptide Ac-WGGGQTAG**$KMe_3$**STG-$NH_2$ (in 10 mM phosphate buffer, pH 8.5), containing a Trp for concentration determination which was separated from the natural sequence by 3 glycine spacers. This peptide also contained Gly8 as opposed to Arg8 which is found in the native sequence. It has been shown that this mutation has minimal affect on the overall binding affinity. The cell was filled with a 0.2 mM solution of either *rac*- or *meso*-**$A_2B$** (in 10 mM phosphate buffer, pH 8.5). A single 0.2 µL peptide aliquot followed by 38 aliquots of 1 µL were titrated into the calorimetric cell every 2.5 minutes

with mechanical stirring. All measurements were carried out at 26°C. The titration data was analyzed using the Origin software by non-linear least squares fitting the heats of binding as a function of the peptide:**$A_2B$** molar ratio to a one site binding model. It was found by ITC that *rac*-**$A_2B$** binds $KMe_3$ with a $K_d$ of 20.0 μM, whereas *meso*-**$A_2B$** binds $KMe_3$ with a $K_d$ of 12.8 μM (Figures 2.47 and 2.48).

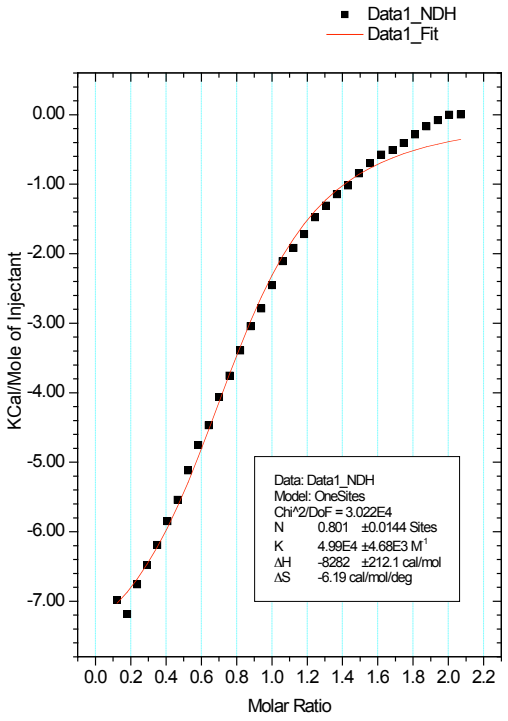



**Figure 2.47.** ITC binding curve from the titration of Ac-WGGGQTAG**$KMe_3$**STG-$NH_2$ into *rac*-**$A_2B$**, giving a $K_d$ of 20.0 μM.

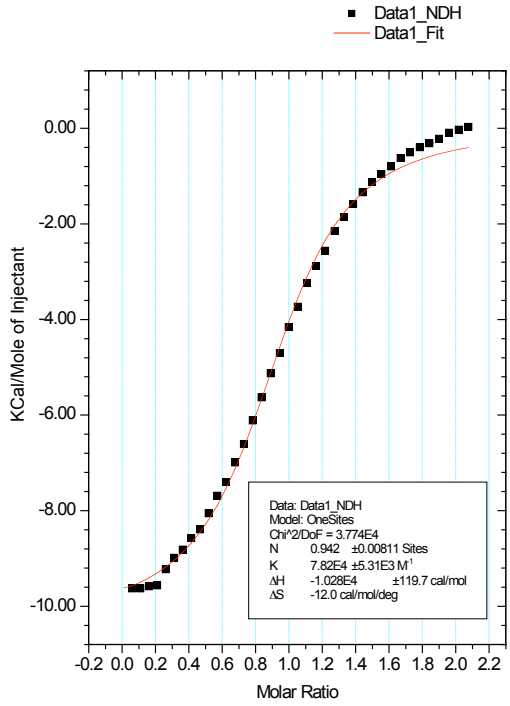


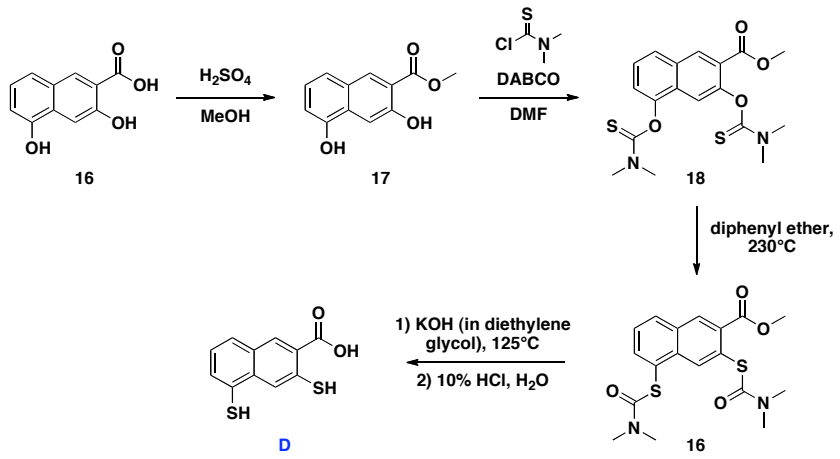

**Figure 2.48.** ITC binding curve from the titration of Ac-WGGGQTAG**KMe$_3$**STG-NH$_2$ into *meso*-**A$_2$B**, giving a $K_d$ of 12.8 μM.

## D. Selective recognition of trimethyllysine and dimethyllysine: *receptor BD$_2$*

### i. Results and discussion

**a. Naphthalene based monomers.** To add a level of diversity to the monomers examined up to this point, we chose to explore dithiol building blocks with a naphthalene core in the investigation of new methylated PTM receptors. Naphthalenes provide a larger aromatic surface area than phenyl groups for interaction with PTMs, and are equally well known for their ability to form cation-π interactions.[13a] Furthermore, the same synthetic steps can be used for the synthesis of such dithiol monomers due to the availability of dihydroxynapthoic acids. This allows for the straightforward synthesis of monomer **D** (Scheme 2.5).[54] The thiols were installed in the 3 and 5 position of the naphthalene ring to favor the formation of cyclic macrocycles over linear oligomers. By

having one thiol on each aromatic ring this also takes advantage of the entire naphthalene surface, whereas if the thiols were placed on a single ring the monomer may be expected to behave similarly to phenyl based monomer **B** or **F**.




**Scheme 2.5.** Synthesis of monomer **D**.

**b. Unbiased library templation studies.** To investigate this naphthalene building block in the context of a DCL, libraries were prepared by mixing equimolar amounts of monomers **B**, **C**, **D**, and **H** (2.5 mM each) in basic solution. The building blocks were dispersed in water and the pH was adjusted to 8.5 with NaOH and HCl. In attempt to identify selective high affinity PTM receptors, all PTM dipeptide templates were screened in eight DCLs at a concentration of 10 mM, while no template was added to the ninth DCL. The libraries were allowed to oxidize in capped HPLC vials at room temperature for about three weeks. All libraries remained completely soluble with no visible precipitation, which is essential. The library compositions were analyzed by LC-MS at various time points, and all library members could be identified by negative electrospray ionization. In some cases, the mass degeneracy of multiple macrocycles prevented the complete characterization of all species without further MS/MS analysis

(for example, **$B_2$CE** and **$D_3$** have identical masses). Upon reaching equilibrium the chromatograms were integrated, and in each case the peak areas were represented as a percentage of the total chromatogram area. This aided in taking into account slight baseline differences from chromatogram to chromatogram.

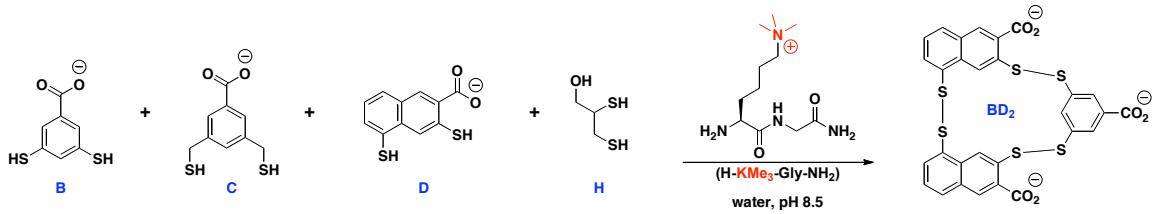



**Figure 2.49.** Dynamic combinatorial library of dithiol monomers resulting in amplification of receptor **$BD_2$** upon addition of a $KMe_3$ guest.

In the absence of guest there are multiple major constituents present, **BC**, **CD**, and **$CD_2$**, as well as numerous smaller peaks corresponding to other macrocyclic library members (Figure 2.50). However, the introduction of a trimethyllysine dipeptide resulted in a clear, significant shift in the library composition in comparison to the untemplated reaction. The amplification of a single receptor in the KMe3 templated library was clearly observed, corresponding to the trimeric macrocycle **$BD_2$** (Figure 2.49). It must be noted that in initial experiments, unacylated dipeptide guests with a free amine on the N-terminus were used for library screens, however it was later shown that amplification of the **$BD_2$** receptor was conserved with the corresponding acylated peptide guests.

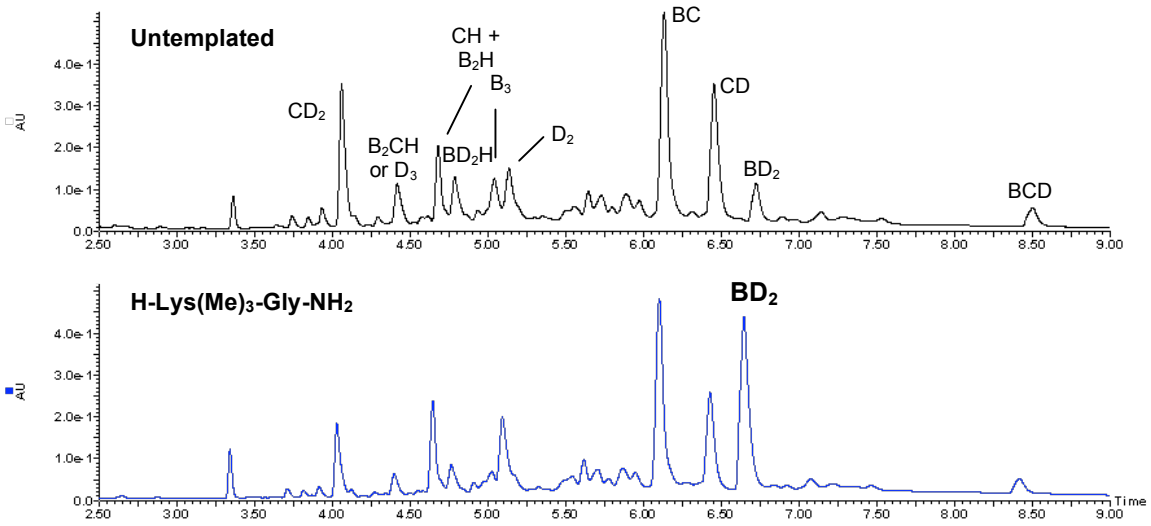



**Figure 2.50.** The analytical HPLC trace at 254 nm of a DCL consisting of monomers **B**, **C**, **D**, and **H** (2.5 mM each), untemplated (black) and in the presence of H-KMe$_3$-G-NH$_2$ (blue).

Next, the selectivity for KMe3 over the lower lysine methylation states and all of the arginine methylation states was examined. The library chromatograms were integrated upon equilibration, and the percent area of the **BD$_2$** peak in each library was calculated by dividing the raw area of the peak by the total area of the chromatogram, not including the area of the peptide guest peak in the case of the templated reactions. While the amplification of **BD$_2$** followed the same general trend as that of **A$_2$B** in that amplification increased with increasing methylation, it was found that **BD$_2$** showed very similar affinity for both KMe2 and KMe3 (Figure 2.51), with a preference for both di- and trimethyllysine over monomethyl lysine. While the comparison of two different libraries is not necessarily an accurate one due to the different equilibria at play, it can be noted that the amplification of **BD$_2$** in the presence of KMe3 is about half that of **A$_2$B**. This may however be due to the higher concentration of **BD$_2$** in the untemplated library in comparison to the concentration *rac*- and *meso*-**A$_2$B** in the untemplated DCLs.

The analysis of the arginine templated libraries revealed amplification of **$BD_2$** in the presence of both sRMe2 and aRMe2, however the extent of amplification is modest in comparison to that of KMe3 and KMe2. Interestingly, **$BD_2$** shows only a minimal preference for aRMe2 over sRMe2 and is essentially unable to discriminate between the two PTMs, however binding to both dimethylated arginines is preferred over monomethylated arginine, following the trend that favorable binding interactions are facilitated by increasing methylation. While the higher lysine methylation states are clearly more favorable than higher arginine methylation states, **$BD_2$** does not appear to be able to distinguish between monomethyllysine and monomethylarginine, which is somewhat surprising due to the significant structural differences between these PTMs.

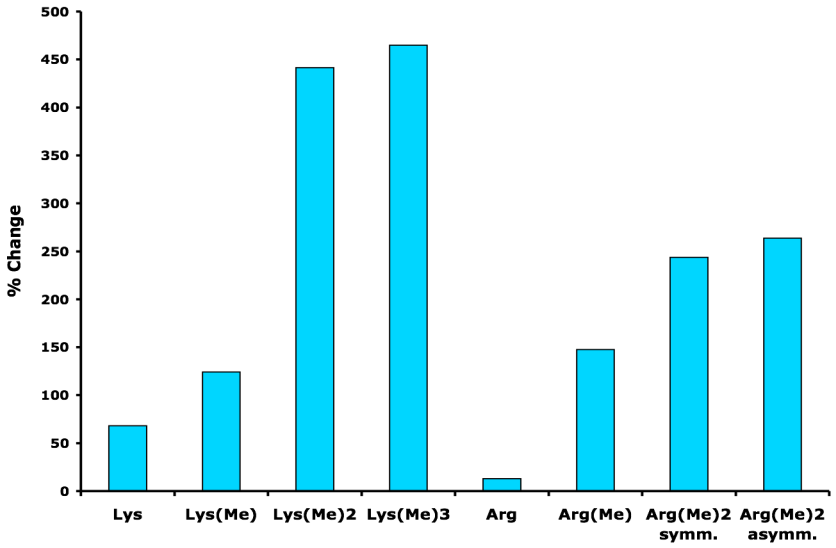



**Figure 2.51.** Amplification of **$BD_2$** in a **BCDH** library templated with all lysine and arginine methylation states, represented as a percent change in peak area (where % change = [(% area templated - % area untemplated) / % area untemplated] × 100).

As in the case with **$A_2B$**, there are likely several factors that result in the observed selectivity of **$BD_2$** for KMe3 and KMe2 over lower lysine and arginine methylation states. The binding pocket may be too large for the lower methylation states, yet it

appears to accommodate both KMe2 and KMe3 equally well. The increased hydrophobicity of the higher methylation states is also more amenable to binding an aromatic pocket. A desolvation penalty may also play a role in the observed selectivity. However, the addition of a third methyl group does not appear to introduce any additional favorable hydrophobic interactions from that of its dimethyl analogue, nor does trimethyllysine's ability to form a stronger cation-π interaction seem to come into play significantly in this case. It is likely that multiple factors, some of which enhance the recognition of KMe2 and others which reduce the affinity for KMe3, work in combination to afford $\mathbf{BD_2}$'s seemingly equal affinity for both PTMs. While $\mathbf{BD_2}$ still shows promise as a small molecule receptor for KMe3 and KMe2, the factors contributing to its recognition of methylated PTMs appear to impart less selectivity than that of $\mathbf{A_2B}$.

**c. Biased library templation studies.** Focusing on the promising recognition of di- and trimethyllysine, additional DCL screens were set up biased towards the formation of $\mathbf{BD_2}$ in which building blocks **B** and **D** were mixed in a 1:2 ratio (7.5 mM total), facilitating the formation of $\mathbf{BD_2}$ in a higher concentration. As expected, a number of macrocycles were generated in the absence of a template, however upon templation with both Lys Me2 and LysMe3 the equilibrium was drastically shifted to form almost exclusively $\mathbf{BD_2}$ (Figure 2.52). This high concentration of $\mathbf{BD_2}$ indicates a quite high thermodynamic stability of this receptor in the presence of both KMe2 and KMe3.

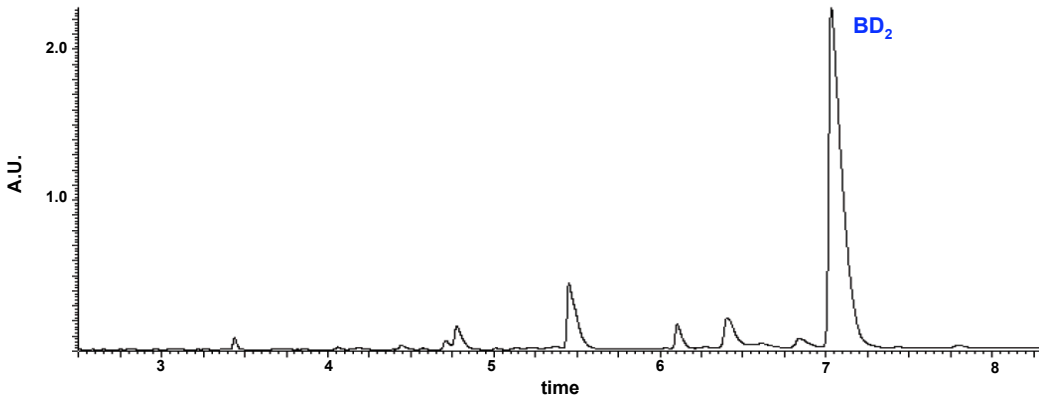



**Figure 2.52.** Part of the analytical HPLC trace at 254 nm of a biased DCL consisting of monomers **D** (5 mM) and **B** (2.5 mM) in the presence of Ac-K$Me_3$-G-$NH_2$.

It is evident that the amplification in this biased library follows a similar trend as that of the prior screen, showing significant selectivity for di- and trimethyl lysine over mono- and unmethylated lysine (Figure 2.53). While the concentration of **$BD_2$** less than doubles in the presence of LysMe and Lys, it increases 400-500 percent when templated with both LysMe2 and LysMe3.

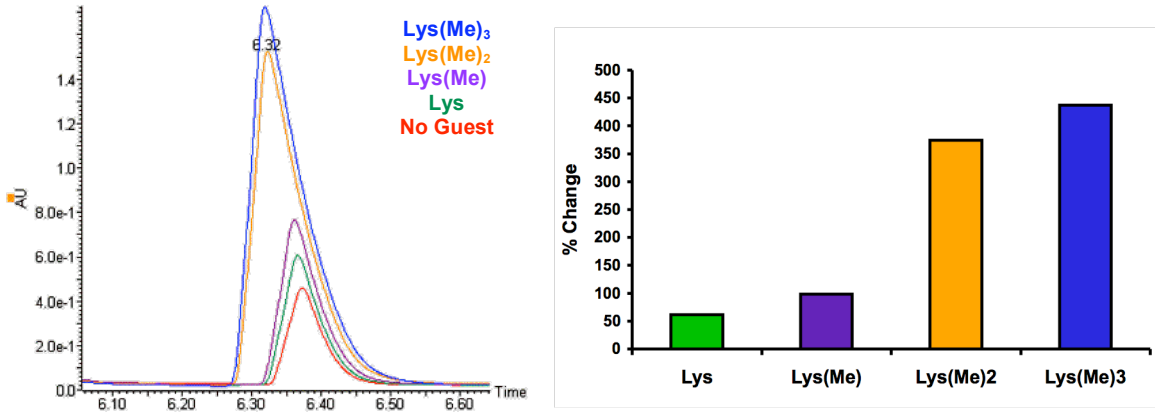



**Figure 2.53.** The extent of amplification of **$BD_2$** with H-K$Me_x$-G-$NH_2$ guests in a biased library relative to the untemplated library (the **$BD_2$** HPLC peak at 254 nm on the left, and the calculated percent change on the right).

**d. Investigation of a histone 3 K9Me3 peptide template.** To ensure that these amplification results were maintained in the context of the naturally methylated histone

peptide and not substantially influenced by surrounding residues, a biased library templated with the 8 residue H3 Lys9Me3 (Figure 2.21) was prepared as described previously and monitored by LC/MS. Significant amplification of **$BD_2$** by H3 Lys9Me3 was observed and determined to be even slightly greater than that of the LysMe3 dipeptide by about 10% (Figure 2.54). This confirms that **$BD_2$** still binds LysMe3 in the context of a histone tail peptide, potentially with even slightly higher affinity. **$BD_2$** amplification in the presence of a histone LysMe2 peptide was not investigated, although based on the above results it is assumed that amplification of LysMe2 in the context of a longer histone tail peptide would be equal to that of the dipeptide.

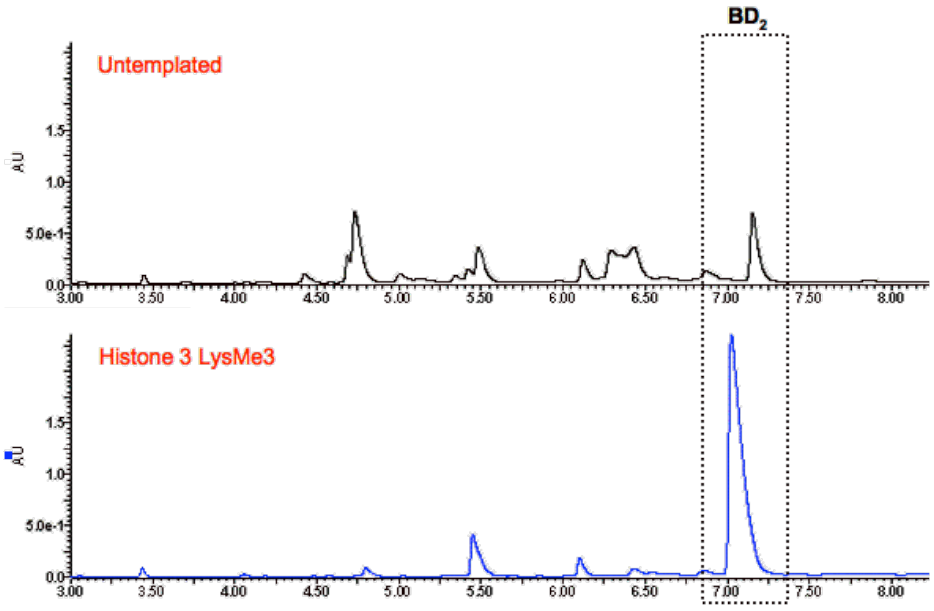



**Figure 2.54.** The analytical HPLC trace at 254 nm of a biased DCL consisting of monomers **D** (5 mM) and **B** (2.5 mM), untemplated (black) and in the presence of H3 LysMe3 (7.5 mM, blue).

**e. Preparative scale biased libraries and isolation of $BD_2$.** To pursue further studies with **$BD_2$** involving both its characterization and application as a di- and trimethyllysine small molecule receptor, preparation and isolation of **$BD_2$** was required.

Large scale biased libraries were prepared with a total volume of 20 mL and a ratio of **B** : **D** : guest of 1 : 2 : 3, with a total monomer concentration (**B** + **D**) of 7.5 mM. Due to the approximately equal affinity of $\mathbf{BD_2}$ for both di- and trimethyllysine, these larger scale libraries were templated with a dimethyl lysine dipeptide due to the simpler synthesis of this peptide. The three solids were combined with the appropriate amount of water, titrated to pH 8.5 with NaOH and HCl, and let stir in a capped scintillation vial for 1-3 weeks before being purified on a semi-preparative HPLC. The purity was confirmed in each case by LC/MS. Relative to the synthesis and purification of *rac*- and *meso*-$\mathbf{A_2B}$, the preparation of $\mathbf{BD_2}$ afforded considerably higher yields, due to both the much higher concentration of $\mathbf{BD_2}$ in the templated library as well as the isolation of only a single isomer.

**f. Characterization of $\mathbf{BD_2}$.** While there is clearly more than one possible structural isomer of $\mathbf{BD_2}$, the amplification of only a single isomer in the presence of methylated PTMs was observed. This is in direct contrast to the non-diastereoselective amplification of both *rac*- and *meso*-$\mathbf{A_2B}$ by LysMe3. This indicates significant structural differences between the three possible $\mathbf{BD_2}$ isomers, two of which are symmetric and one of which is asymmetric (Figure 2.55). More interesting is the fact that one isomer is significantly more thermodynamically stable than the others, as a statistical ratio of the three isomers is not observed in the untemplated library. Instead, only one peak corresponding to the mass of $\mathbf{BD_2}$ is observed in a significant quantity, and this may indeed be the reason that only a single $\mathbf{BD_2}$ macrocycle is amplified.

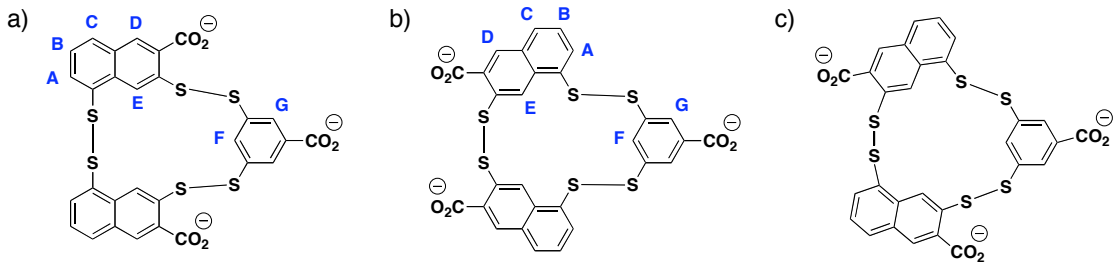



**Figure 2.55.** The potential structural isomers of **$BD_2$**, two of which are symmetric (a and b) while the third is asymmetric (c).

It was initially suspected that the three isomers were co-eluting during library analysis due to their similarities in structure, polarity, and hydrophobicity. However during efforts to improve the chromatographic separation, there was no evidence of a second **$BD_2$** isomer. The purified receptor was then analyzed by NMR to gain further insight as to it's purity and structure. The 1D $^1$H NMR of the receptor revealed 6 sharp proton signals, with one very broad signal. The simplicity of the NMR spectra indicated that the purified receptor was not only a single isomer, but one of the two symmetric isomers, as 13 distinct signals would be expected for the asymmetric isomer.

All protons were assigned through 2D TOSCY and NOESY experiments, with proton F determined to be the broadened signal. While the signals for protons A-E and G get sharper above room temperature, the broadening of proton F increases. However, at 5 °C the signal for proton F gets slightly sharper. It is thus predicted that this proton typically undergoes a significant amount of rotation resulting in a broadened signal, and that this rotation is restricted to some degree at lower temperatures. While 1D experiments were successfully run above room temperature at both 40 °C and 60 °C, it must be noted that longer 2D experiments at 60 °C provided sufficient time and energy for the disulfide bonds to break, resulting in re-equilibration of the library as determined by subsequent

LC/MS analysis of the NMR sample. The exact temperature at which this re-equilibration begins to occur was not investigated further. 2D NOESY experiments did not provide sufficient evidence to determine which of the two **$BD_2$** symmetric isomers was being amplified (Figure 2.55a and 2.55b), and while this information is interesting and would provide a more complete characterization, it does not largely impact the significance of this work.

**g. NMR analysis of trimethyllysine binding to $BD_2$.** In addition to the characterization of **$BD_2$**, NMR was also used to investigate the recognition of a trimethyllysine dipeptide by **$BD_2$**. Three samples were prepared and analyzed under identical conditions: a) **$BD_2$** only, b) H-K$Me_3$-G-$NH_2$ dipeptide only, and c) H-K$Me_3$-G-$NH_2$ bound to **$BD_2$**. Upon the introduction of **$BD_2$**, significant broadening of all KMe3 proton signals was observed, indicating a change in the peptide's environment. A similar broadening of the **$BD_2$** proton signals was also observed upon binding. Furthermore, all lysine side chain protons were upfield shifted upon binding, with the extent of shifting increasing moving away from the peptide backbone and closer to the site of methylation. The larger shifts are indicative of a greater degree of interaction with **$BD_2$**. A similar extent of shifting was observed for the δ, ε, and methyl protons, indicating that the receptor is not only interacting with the methylammonium group, but is also in close contact with both the δ and ε methylene protons (Figure 2.56). This is not entirely surprising, as it has previously been shown that a LysMe3 residue in a β-hairpin peptide interacts preferentially with the face of a tryptophan ring through both its polarized ε methylene and its methylammonium group.[55] It is possible that the naphthalene rings of

**BD$_2$** are interacting with the KMe3 sidechain in a similar fashion as the indole ring of Trp.

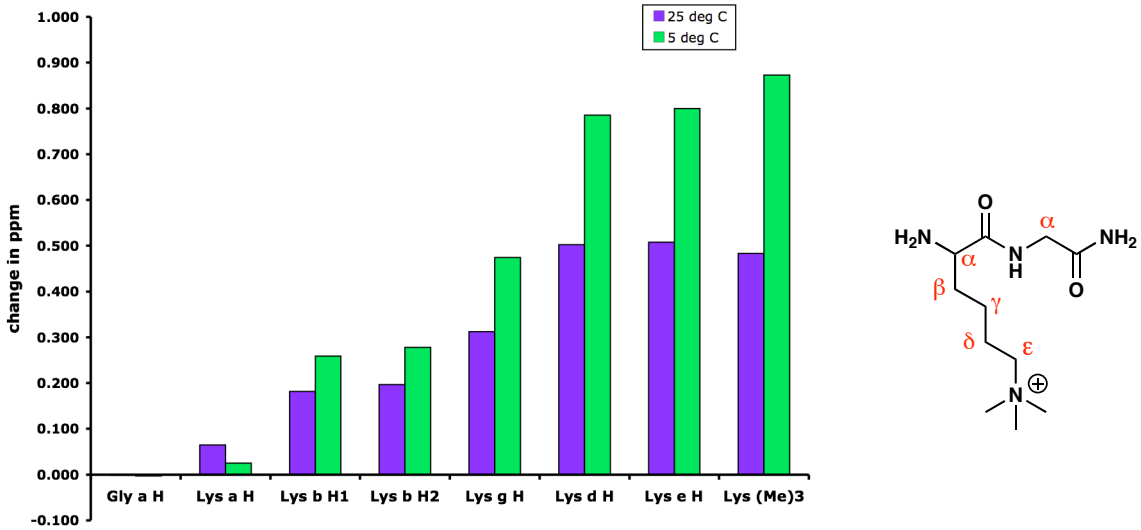



**Figure 2.56.** Upfield side-chain chemical shifts of H-KMe$_3$-G-NH$_2$ upon binding to **BD$_2$** at both 25 °C and 5 °C. Conditions: 50 mM borate pD 9.0 with 10% acetonitrile-d$_3$, referenced to DSS.

Both the temperature and the solvent system seem to have a significant affect on the extent of LysMe3 upfield shifting, and therefore the degree of interaction of the two molecules. The LysMe3 protons are all upfield shifted more drastically at 5 °C. This may indicate that the guest is in a bound state a larger percent of the time at 5 °C, assuming the conversion between the bound and unbound state is fast on the NMR timescale. In regards to the solvent system, the amount of an added co-solvent appears to influence the degree of upfield shifting. In figure 2.56 the LysMe3 upfield shifts upon binding in 10% acetonitrile-d$_3$ in borate buffer are presented. If the ratio of acetonitrile to aqueous buffer is increased to 1:2 only minimal upfield shifting of the LysMe3 protons is observed. However, in aqueous buffer only, a 0.86 ppm upfield shift of the methyl groups was observed at 25 °C, in comparison to about 0.5 ppm at 25 °C in 10%

acetonitrile. Based on such observations, it is evident that the introduction of an organic solvent has a negative influence on binding, as the driving force for KMe3 binding to **$BD_2$** is likely less when the unbound KMe3 exists in a more hydrophobic environment.

**h. $BD_2$ aggregation by NMR and fluorescence.** Based on evidence that **$BD_2$** functions as a methylated lysine small molecule receptor, we sought to investigate the binding affinity of **$BD_2$** to LysMe3 and LysMe2. This was first attempted by an NMR titration, and while an accurate binding constant was not determined, this study contributed to our overall understanding of this PTM receptor. With the LysMe3 dipeptide concentration constant, **$BD_2$** was titrated into solution to give concentrations ranging from 20 μM to 2.8 mM. While it was anticipated that plotting the upfield chemical shift of the lysine methyl groups against the concentration of **$BD_2$** would provide an appropriate binding curve, it was the **$BD_2$** spectra at varying concentrations that proved to be more informative.

As the concentration of **$BD_2$** increased, even well beyond the point where **$BD_2$** was saturated with LysMe3, the proton signals continued to shift upfield indicating some degree of aggregation of **$BD_2$** (Figure 2.57). The stacking of naphthalene moieties is a common phenomenon that is highly concentration dependent, which explains the more drastic shifts observed with larger jumps in concentration. The **$BD_2$** proton signals remain well resolved at low concentrations, however at higher concentrations a significant broadening effect is observed, likely due to a much greater extent of intermolecular stacking between naphthalene units. Interestingly, while protons A-E were continually upfield shifted with increasing **$BD_2$** concentrations, the signal for proton G was not shifted throughout the titration. It is therefore thought that the phenyl ring of

monomer **B** does not participate in any intermolecular interactions or contribute significantly to aggregation. Based on the apparent self-association of this receptor, the determination of an accurate binding constant by NMR and other methods has proved to be challenging, and while **$BD_2$** undoubtedly binds LysMe2 and LysMe3 selectively over other lysine and arginine methylation states, the absolute affinity and selectivity has yet to be quantified.

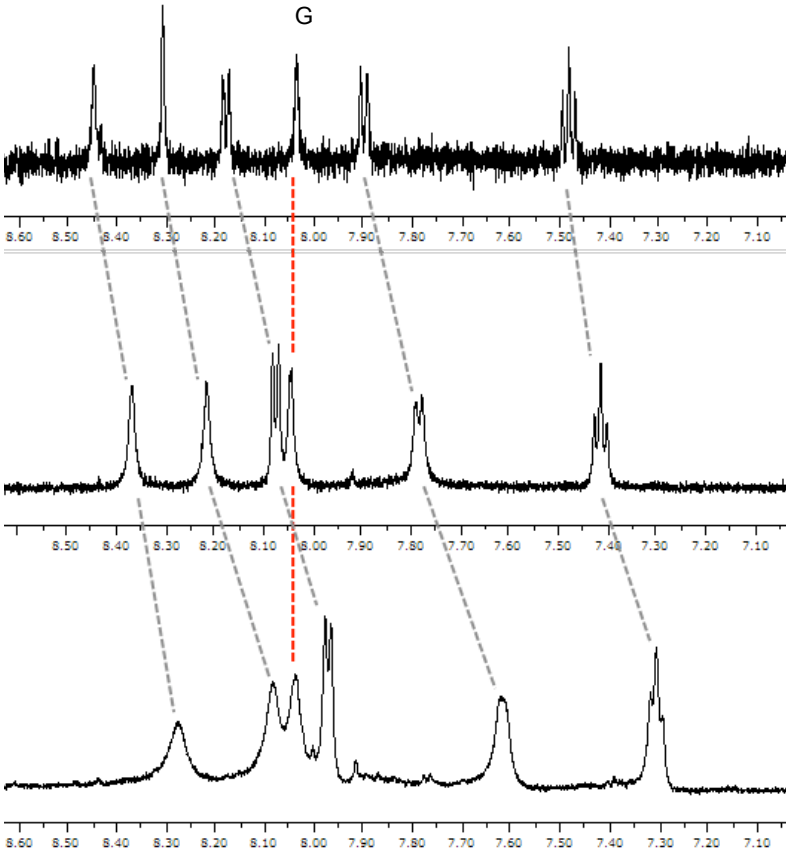



**Figure 2.57.** $^1$H NMR of **$BD_2$** at varying concentrations (100 μM top, 600 μM middle, and 2.0 mM bottom) in the presence of H-K$Me_3$-G-$NH_2$ (100 μM). The chemical shift of proton G (red line) remains unchanged, whereas protons A-E (gray lines) shift upfield.

To further investigate this aggregation phenomenon, the fluorescence properties of **$BD_2$** were explored. Naphthalene molecules are known to fluoresce from about 300-330

nm; however, any aggregation of these molecules would be expected to result in a "self-quenching" of the naphthalene fluorescence to some degree. Similarly, naphthalene is capable of exhibiting excimer fluorescence at longer wavelength regions around 390 nm, in addition to the monomer fluorescence at a shorter wavelength.[56] Excimers are short-lived dimeric species formed from two molecules, at least one of which is in an excited state. Generally as the monomer fluorescence decreases, the excimer fluorescence increases, and vise versa. Upon excitation of a high concentration solution of **$BD_2$** (1 mM) at 240 nm, there was no observable fluorescence of the individual receptors at shorter wavelengths, as it is presumably all quenched as a result of **$BD_2$** aggregation (Figure 2.58). Interestingly, at this concentration fluorescence is observed at higher wavelengths, indicative of stacking of the naphthalene moieties, and thus the formation of excimers.

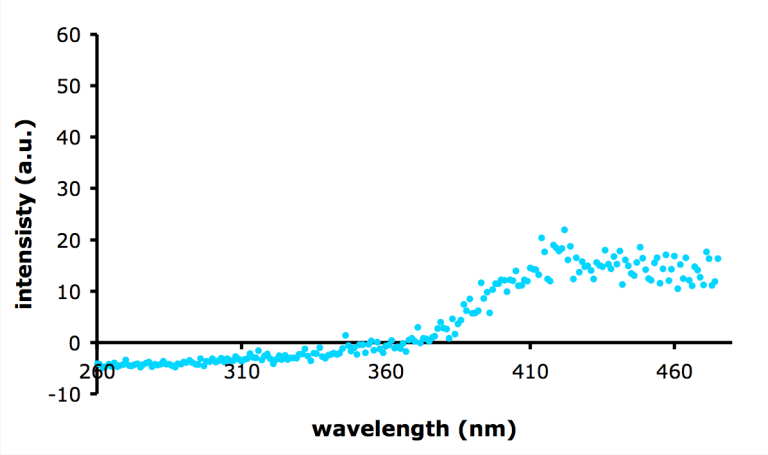



**Figure 2.58.** Fluorescence spectra of a 1 mM solution of **$BD_2$** upon excitation at 240 nm.

At lower concentrations of **$BD_2$** (20 μM), some degree of fluorescence is observed in the shorter wavelength region around 310 nm, as aggregation is likely occurring to a much lesser degree at this concentration (Figure 2.59). Upon the addition of an excess of

a KMe3 dipeptide, this fluorescence increases. This indicates that **$BD_2$** is indeed interacting with the trimethyllysine peptide, and in doing so is breaking up some of the **$BD_2$** aggregates and decreasing the degree of fluorescence quenching. Overall, the **$BD_2$** fluorescence spectra support the NMR data, confirming a significant degree of aggregation of **$BD_2$**.

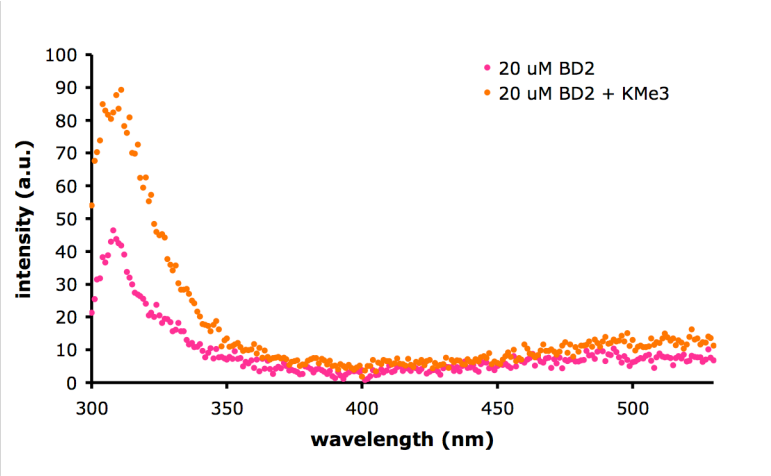



**Figure 2.59.** Fluorescence spectra of a 20 μM solution of **$BD_2$** upon excitation at 278 nm (pink), and in the presence of 60 μM KMe3 (orange).

**i. Structure-function studies.** To evaluate the specificity of the interaction between **$BD_2$** and KMe3 and KMe2 in comparison to other structurally similar receptors, simple systematic structural variations were made to both monomers **B** and **D**. How these changes in monomer structure influence the enhancement of **$BD_2$** and thus the function of these receptors in various libraries templated with methylated PTMs was then investigated. One structural isomer of each monomer was investigated. Monomer **E** (Figure 2.12) was investigated as a variation of **D**, where the positions of the thiols were changed from the 3- and 5-positions of the naphthalene ring to the 3- and 7-positions. In monomer **E** the thiols are now placed further apart and on opposite sides of the

naphthalene moiety in comparison to **D**. The synthesis of monomer **E** is identical to that of **D**, starting with 3,7-dihydroxynapthoic acid. In addition, monomer **B'**, where the thiols are placed in the 2- and 6-positions of the ring relative to the carboxylate, was investigated as a variation of monomer **B**. By screening monomers **E** and **B'** in DCLs against methylated PTM guests, we can gain insight into how these subtle variation influences binding.

In the investigation of monomer **E**, all methylated lysine and arginine dipeptide guests were screened in unbiased DCLs containing monomers **B**, **C**, **E**, and **H** (2.5 mM each). After over a week, no significant amplification was observed in any of the templated libraries, with macrocycle **BC** as the dominant species in each case. Interestingly, there was also evidence of the formation of linear oligomer chains of **E**. Due to the spacing of the thiols in **E** it is not entirely surprising that monomer **D** lends itself more towards the formation of cyclic macrocycles, whereas **E** favors linear oligomers. This reiterates the importance of the placement of reactive functional groups within monomers, as functional groups closer together tend to favor cycle macrocycles whereas those placed on opposite sides are more likely to form linear oligomer chains in the absence of binding interactions. It is clear that despite their structural similarities, replacing monomer **D** with **E** has a drastic effect on both the composition of the DCLs formed as well as the templation of a potential PTM receptor. While a DCL containing monomers **B**, **C**, **D**, and **H** resulted in a range of amplification in the presence of all methylated PTMs, in a similar DCL containing monomers **B**, **C**, **E**, and **H** no significant amplification was observed.

In the investigation of monomer **B'**, biased analytical DCLs were setup by mixing building blocks **D** and **B'** in a 2:1 ratio (7.5 mM total), and the recognition of the different methylated PTMs by macrocycles in the DCL was evaluated. As in the case of the **$A_2$B'** DCLs, the library composition of these libraries was largely affected by this structural change. Very minor amplification of peaks corresponding to **B'$D_3$**, **B'$_2D_2$**, and **B'$_4D_2$** was observed in the presence of both symmetric and asymmetric ArgMe2, however biasing the libraries towards these receptors did not enhance amplification, indicating extremely weak binding.

In contrast to the observed amplification of **$BD_2$**, there does not appear to be favorable recognition of either LysMe2 or LysMe3 by **B'$D_2$** (Figure 2.60). This may be partially due to the fact that **B'$D_2$** is by far the dominant macrocycle in both templated and untemplated libraries, indicating a high thermodynamic stability of this macrocycle prior to binding. The thermodynamic stability of **B'$D_2$** is significantly higher than that of **$BD_2$**, as **$BD_2$** exists in much lower concentrations in untemplated libraries, again displaying the drastic change in library composition upon the introduction of slight structural modifications. In attempt to reduce the effective concentration of **B'$D_2$** in the untemplated library, the ratio of **B'**:**D** was varied systematically from 1:2 to 1:1, 2:1, and 4:1 to bias it away from **B'$D_2$**. While increasing the ratio of **B'**:**D** reduced the concentration of **B'$D_2$** to some degree, this was accompanied by a concomitant increase in the concentration of macrocycles **B'$_5$ - B'$_8$** with still no significant **B'$D_2$** amplification.

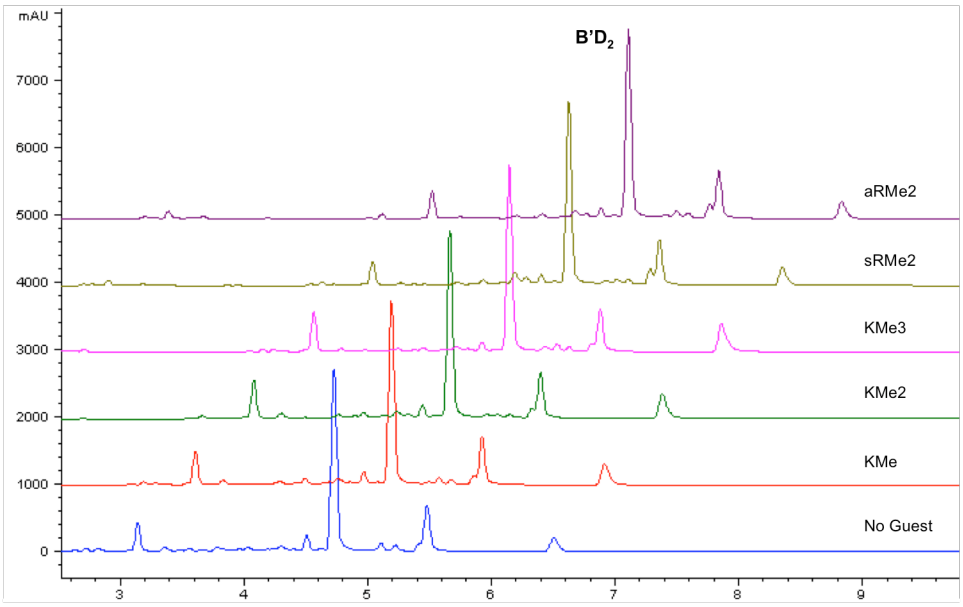



**Figure 2.60.** Part of the analytical HPLC trace at 254 nm of a biased DCL consisting of monomers **D** (5 mM) and **B'** (2.5 mM), untemplated and in the presence of methylated lysine and arginine (7.5 mM) dipeptides.

### ii. Conclusions

In conclusion, we report the identification of a synthetic receptor, $\mathbf{BD_2}$, which selectively recognizes both KMe2 and KMe3 over mono- and umethylated lysine, as well as all arginine methylation states. This small molecule receptor was identified through dynamic combinatorial chemistry, displaying the novelty and utility of this technique in the discovery of synthetic receptors for post-translational modifications. While $\mathbf{BD_2}$ is selective for both KMe2 and KMe3 PTMs, it is likely that similar factors contribute to the selectivity of this naphthalene based receptor as those contributing to the selective recognition of KMe3 by $\mathbf{A_2B}$. The observed selectivity for higher methylation states can be largely attributed to differences in the magnitude of the cation-π interactions, as well as differences in size and desolvation penalty. The difference in amplification observed when **B** and **D** are replaced with structural variants is interesting, and demonstrates that

subtle structural changes can have a significant effect on recognition and function. Although **$BD_2$** shows promise as a small molecule receptor for KMe2 and KMe3, an absolute affinity for either PTM has not been determined due to the observed aggregation of this receptor. However, it is conceivable that this could indeed be achieved with further studies if the concentration of **$BD_2$** remains below that of where aggregation begins to occur. Similarly, this receptor also has the potential to be useful in a variety of applications in the investigation of unknown protein lysine methylation states when concentrations are required where aggregation is not problematic.

### iii. Experimental Section.

**a. Dynamic combinatorial library screens.** Library screens were conducted similarly to those described previously. The relevant building blocks were individually dissolved in water with sufficient 1.0 M aqueous NaOH to bring them into solution, and the pH of each solution was then adjusted to 8.5. In the unbiased **BCDH** libraries, aliquots of each monomer solution were combined in a 2 mL LC-MS vial to reach a final concentration of 2.5 mM of each monomer (10 mM total). In the biased **$BD_2$** libraries, aliquots of each monomer solution were combined to reach a final concentration of 5 mM **D** and 2.5 mM of **B** (7.5 mM total). When necessary, aliquots of peptide guests dissolved in water were added to the reactions to reach a final concentration of either 10 mM for the **BCDH** libraries or 7.5 mM for the **$BD_2$** libraries. The vials were capped and analyzed at various time points.

**b. Analytical LC/MS.** LC-MS was carried out on either a Waters Acquity LC-MSD system or a Agilent Rapid Resolution LC-MSD system, equipped with an online

degasser, binary pump, autosampler, heated column compartment, and diode array detector. All separations were performed using 5 mM $NH_4OAc$ $H_2O$-acetonitrile gradients at pH 5. The MS was performed using a single quad mass spectrometer and all peaks were identified by negative electrospray ionization. The reactions were monitored weekly (2 μL or 10 μL injections depending on total monomer concentration) until equilibrium was reached after about 3 weeks. On the Waters Acquity system the chromatography of the **BCDH** and **$BD_2$** DCLs was carried out at 40°C with gradient D (Table 2.5). On the Agilient Rapid Resolution system the chromatography of the **$BD_2$** DCLs was carried out at 50°C with gradient E, and the chromatography of the **$B'D_2$** DCLs was carried out with gradient F using a gradient temperature, going from 50°C to 40°C, left to right (Halo C18 column, 4.6 × 100 mm, 2.7 micron). The peak areas were integrated at 254 nm.

**Table 2.5.** Analytical LC methods use to analyze **BCDH** and **$BD_2$** DCC libraries on the Waters Acquity system (Method D) and the **$BD_2$** libraries (Method E) and **$B'D_2$** libraries (Method F) on the Agilent Rapid Resolution

| **Gradient D** (Waters Acquity) | | |
|---|---|---|
| **Time (min)** | **%B** | **Flow Rate (mL/min)** |
| 0.00 | 0.0 | 0.4 |
| 1.00 | 20.0 | 0.4 |
| 3.00 | 29.0 | 0.4 |
| 9.00 | 44.0 | 0.4 |
| 9.10 | 100.0 | 0.4 |
| 10.00 | 100.0 | 0.4 |
| 10.0 | 0.0 | 0.4 |
| 12.00 | 0.0 | 0.4 |

| **Gradient E** (Agilent Rapid Resolution) | | |
|---|---|---|
| **Time (min)** | **%B** | **Flow Rate (mL/min)** |
| 0.00 | 3.0 | 1.0 |
| 2.00 | 30.0 | 1.0 |
| 9.00 | 48.0 | 1.0 |
| 10.00 | 100.0 | 1.0 |
| 12.00 | 100.0 | 1.0 |
| 12.50 | 3.0 | 1.0 |
| 16.00 | 3.0 | 1.0 |

| **Gradient F** (Agilent Rapid Resolution) | | |
|---|---|---|
| **Time (min)** | **%B** | **Flow Rate (mL/min)** |
| 0.00 | 3.0 | 1.0 |
| 3.00 | 30.0 | 1.0 |
| 8.25 | 30.0 | 1.0 |
| 10.00 | 40.0 | 1.0 |
| 10.10 | 100.0 | 1.0 |
| 12.00 | 100.0 | 1.0 |
| 12.10 | 3.0 | 1.0 |
| 16.00 | 3.0 | 1.0 |

**c. Synthesis and isolation of $BD_2$.** Biased libraries were prepared on a 19.3 mL scale (**B**: 9.0 mg, 0.048 mmol; **D**: 22.8 mg, 0.097 mmol) templated with a H-KMe2-G-$NH_2$ (33.4 mg, 0.145 mmol). Upon equilibration the libraries were neutralized and the receptors were isolated by semi-preparative HPLC (Figure 2.61). Approximately 0.5 mL injections were chromatographed using standard peptide synthesis mobile phases A and B (0-55% B 0-5 min, then 55-100% B 5-30 min) with a flow rate of 4.0 mL/min. The **$BD_2$** peak was collected and analyzed for purity by analytical LC-MS (Figure 2.62). Purified **$BD_2$** was lyophilized to powder and stored under nitrogen.

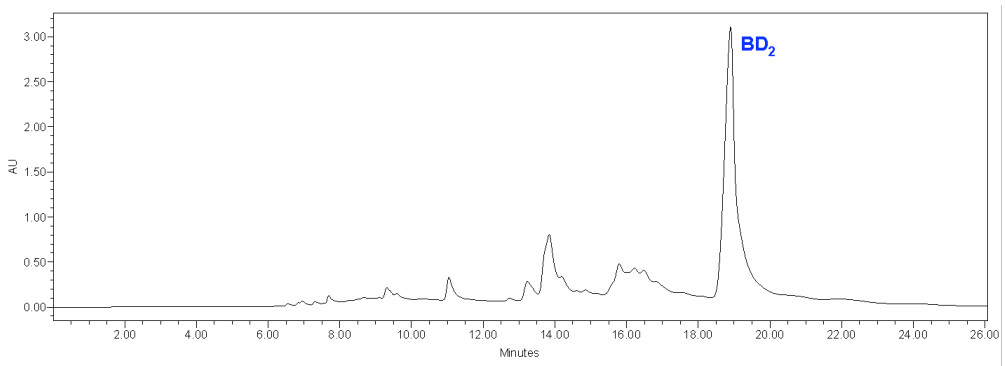



**Figure 2.61.** Semi-preparative HPLC trace of a **$BD_2$** biased library.

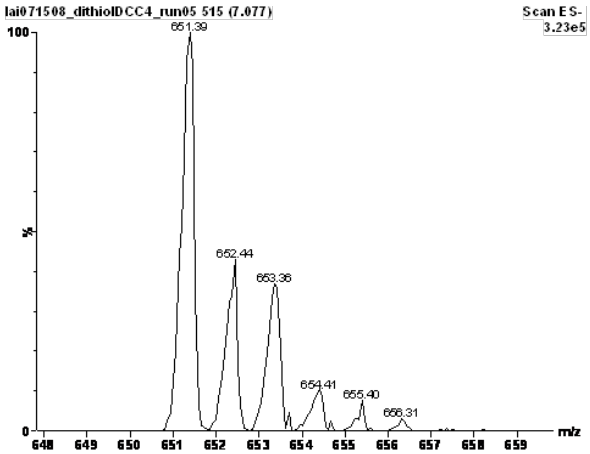
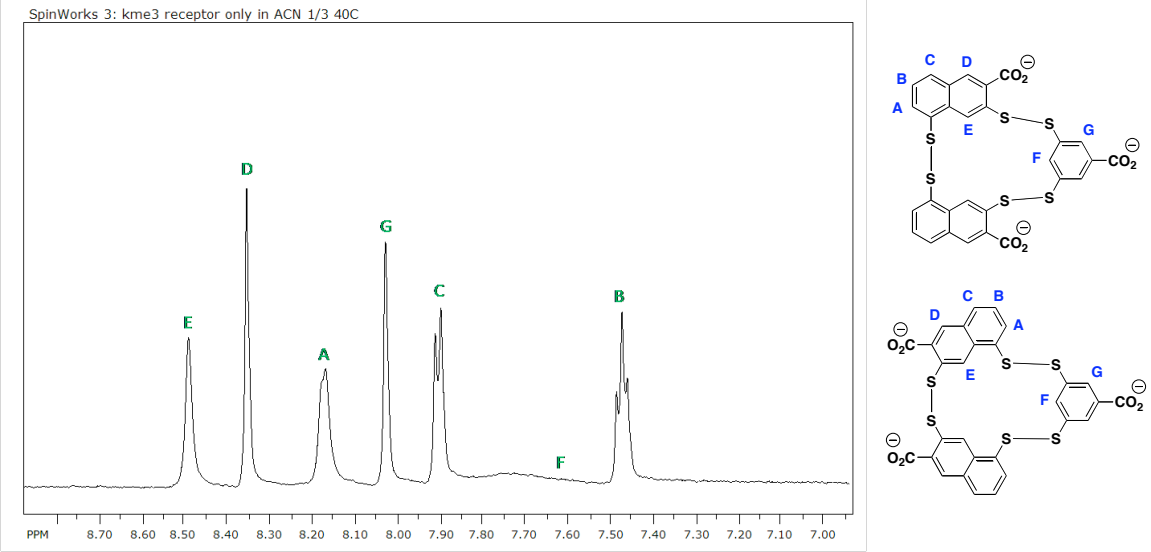



**Figure 2.62.** Mass spectra of purified **$BD_2$** (-ESI).

**d. NMR spectroscopy.** All NMR samples were prepared and analyzed on a Varian Inova 600-MHz instrument similarly to those described previously, unless noted otherwise. For characterization, a **$BD_2$** sample (approximately 2 mM) was prepared in 67% 50 mM borate buffer at approximately pD 9 containing 0.05 mM DSS and 33% acetonitrile-$d_3$. There appeared to be significantly less aggregation at such high concentrations with the addition of a co-solvent such as acetonitrile. 1D spectra were collected at 5 °C, 25 °C, and 40 °C using 32K data points and 16-32 scans with a 1-3 second presaturation. All 2D NMR experiments used pulse sequences from the Chempack software including TOSCY and NOESY. TOSCY experiments were taken

with 16 scans in the first dimension and 128 scans in the second dimension. NOESY experiments were taken with 32 scans in the first dimension and 256 scans in the second dimension, with a mixing time of 0.8 seconds. A number of cross peaks were identified in both TOSCY and NOESY experiments allowing for the assignment of all protons (Figure 2.66).




$^1$H NMR (600 MHz): δ = 7.47 (t, 2 H), 7.58 (s, 1H), 7.91 (d, 2H), 8.03 (s, 2H), 8.17 (d, 2H), 8.36 (s, 2H), 8.49 (s, 2H).

**Figure 2.63.** $^1$H NMR of **BD$_2$** at 40 °C in borate buffer with 1/3 acetonitrile-d$_3$ (left) and the assigned protons in both possible symmetric isomers (right).

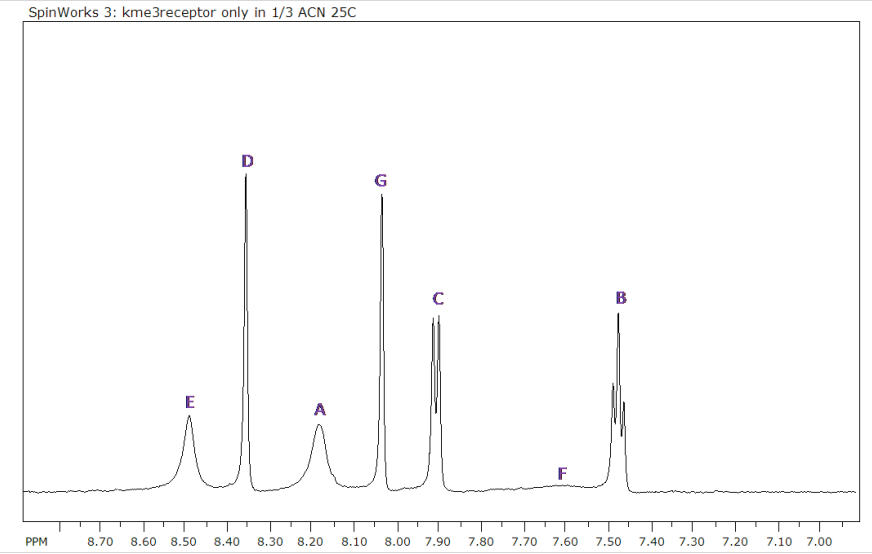



**Figure 2.64.** $^{1}$H NMR of **BD$_2$** at 25 °C in borate buffer with 1/3 acetonitrile-d$_3$.

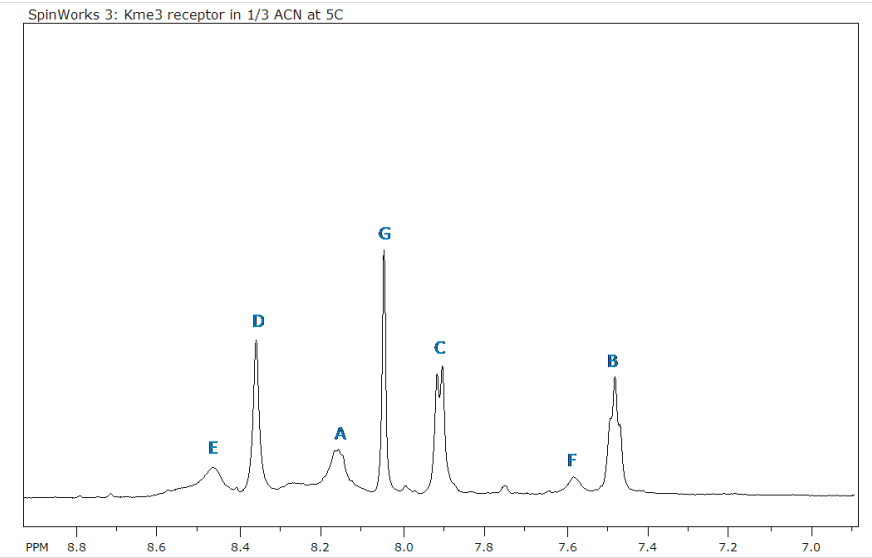



**Figure 2.65.** $^{1}$H NMR of **BD$_2$** at 5 °C in borate buffer with 1/3 acetonitrile-d$_3$.

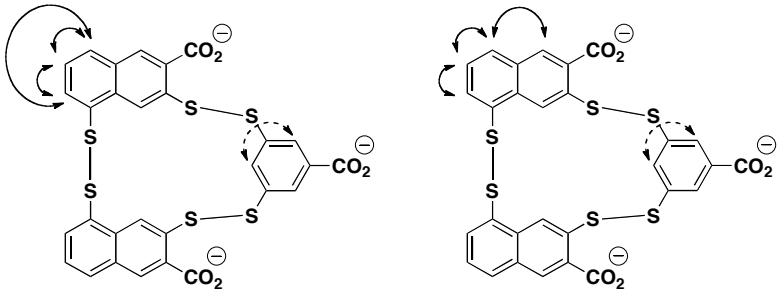


**Figure 2.66. $BD_2$** TOSCY cross peaks (left) and NOESY cross peaks (right). The cross peak between protons F and G is weaker in both cases, as indicated by a dashed arrow.

To examine LysMe3 binding to **$BD_2$,** 1D NMR spectra of three samples were collected as described above. Two samples, one containing **$BD_2$** and the other H-$KMe_3$-G-$NH_2$, were prepared in 50 mM borate pD 9.0 with 10% acetonitrile-$d_3$, each with a concentration of approximately 4 mM (600 μl). Upon collection of 1D spectra for both samples at 25 °C and 5 °C, 10 μl of a 261 mM H-$KMe_3$-G-$NH_2$ stock solution was added to the **$BD_2$** sample to give approximately equal concentrations of both receptor and guest, and new spectra were collected with 64 scans. Due to a significant number of buffer impurities, a spectra of the buffer alone was collected to ensure everything in the spectra could be properly accounted for.

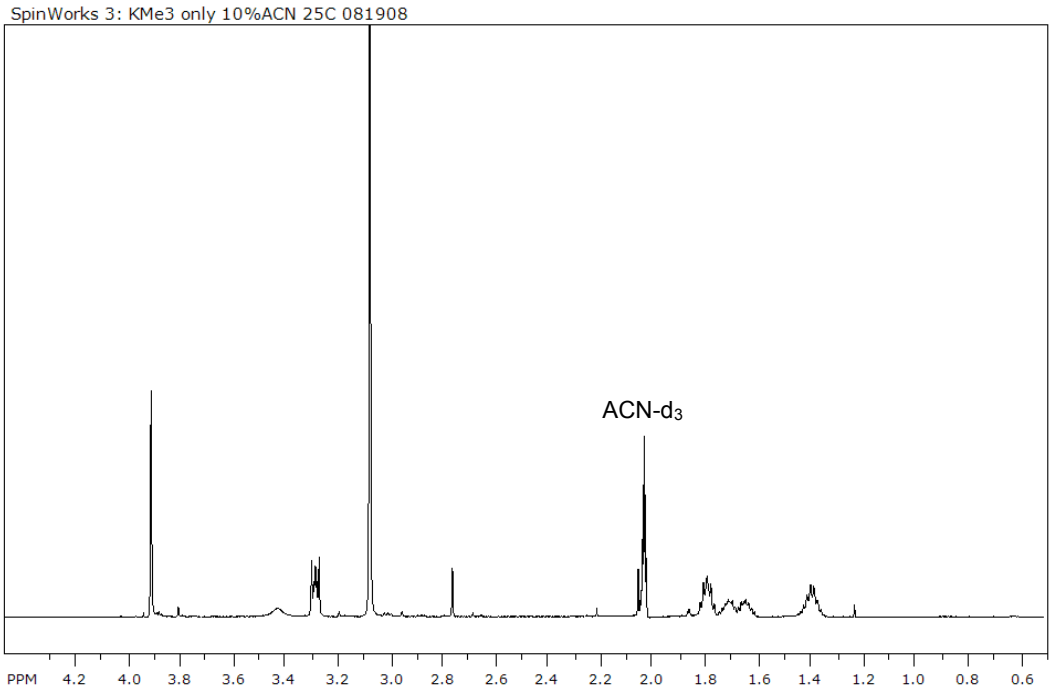



**Figure 2.67.** $^{1}$H NMR of H-KMe$_3$-G-NH$_2$ at 25 °C.

**Table 2.6.** Proton chemical shift assignments for $^{1}$H NMR of H-KMe$_3$-G-NH$_2$ at 25 °C.

| Residue | α | β | γ | δ | ε | Me |
|---|---|---|---|---|---|---|
| **KMe$_3$** | 3.435 | 1.706, 1.646 | 1.390 | 1.792 | 3.289 | 3.080 |
| **G** | 3.917 | | | | | |

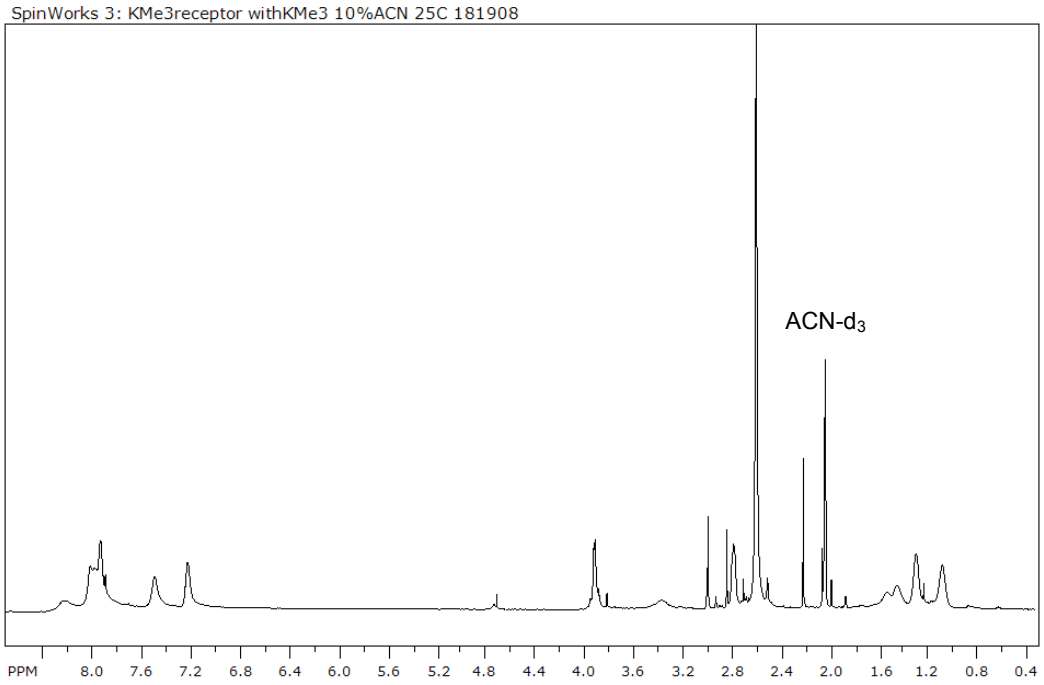



**Figure 2.68.** $^{1}$H NMR of H-KMe$_3$-G-NH$_2$ with **BD$_2$** at 25 °C.

**Table 2.7.** Proton chemical shift assignments for $^{1}$H NMR of H-KMe$_3$-G-NH$_2$ bound to **BD$_2$** at 25 °C.

| Residue | α | β | γ | δ | ε | Me |
|---|---|---|---|---|---|---|
| **KMe$_3$** | 3.370 | 1.524, 1.449 | 1.078 | 1.290 | 2.781 | 2.597 |
| **G** | 3.918 | | | | | |

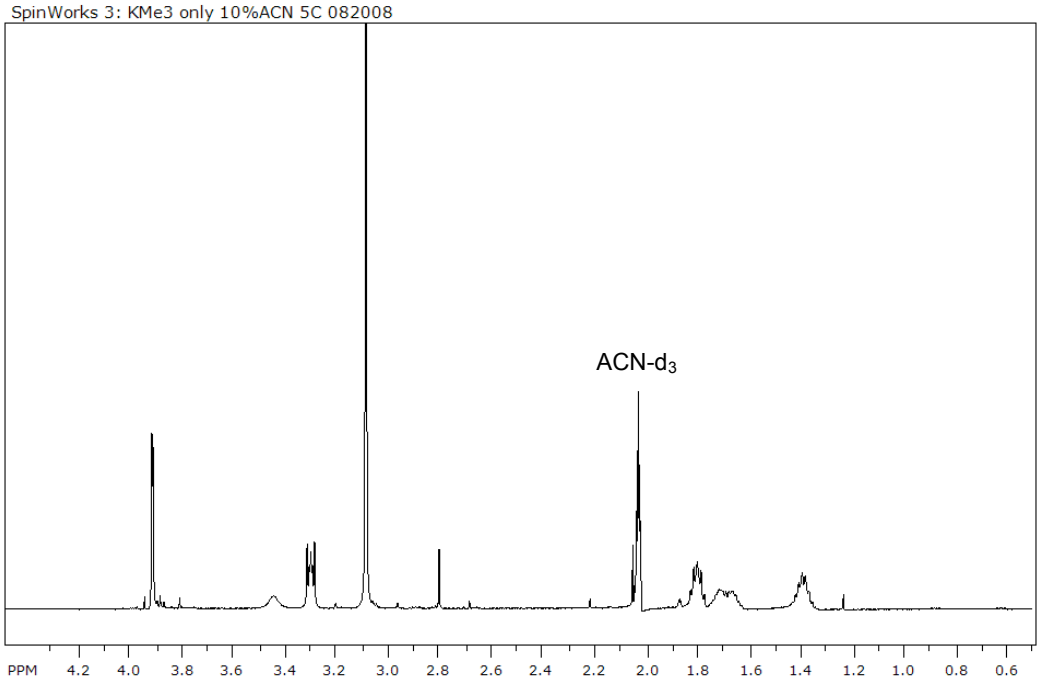



**Figure 2.69.** $^{1}$H NMR of H-KMe$_3$-G-NH$_2$ at 5 °C.

**Table 2.8.** Proton chemical shift assignments for $^{1}$H NMR of H-KMe$_3$-G-NH$_2$ at 5 °C.

| Residue | α | β | γ | δ | ε | Me |
|---|---|---|---|---|---|---|
| **KMe$_3$** | 3.446 | 1.704, 1.660 | 1.384 | 1.798 | 3.302 | 3.088 |
| **G** | 3.918 | | | | | |

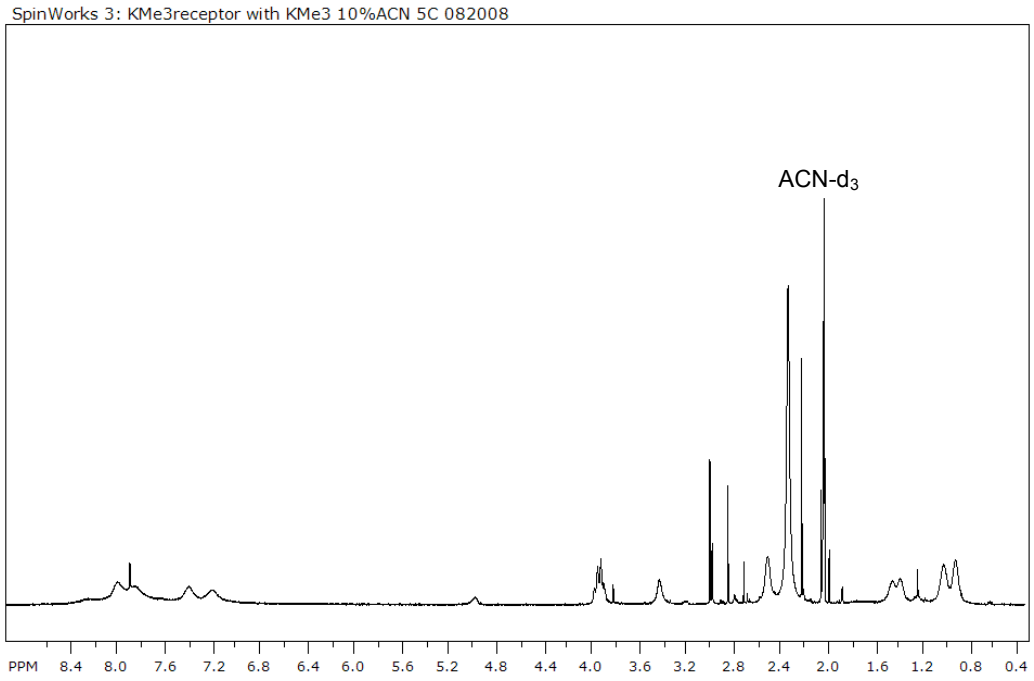



**Figure 2.70.** $^{1}$H NMR of H-KMe$_3$-G-NH$_2$ with **BD$_2$** at 5 °C.

**Table 2.9.** Proton chemical shift assignments for $^{1}$H NMR of H-KMe$_3$-G-NH$_2$ bound to **BD$_2$** at 5 °C.

| Residue | α | β | γ | δ | ε | Me |
|---|---|---|---|---|---|---|
| **KMe$_3$** | 3.421 | 1.445, 1.382 | 0.910 | 1.013 | 2.502 | 2.331 |
| **G** | 3.920 | | | | | |

For the NMR titration experiment, a 100 μM sample of H-KMe$_3$-G-NH$_2$ was prepared in 50 mM borate pD 9.0 with 10% acetonitrile-d$_3$. A 12 mM stock solution of **BD$_2$** was also prepared in 50 mM borate pD 9.0 with 10% acetonitrile-d$_3$, and 1 to 30 μl aliquots were titrated into the KMe3 sample. 1D NMR spectra were collected before the addition of **BD$_2$** and with each successive addition at 25 °C using 64 scans. The concentration of **BD$_2$** in each spectrum was calculated by taking into account the sample volume increase with each addition. The spectra at **BD$_2$** concentrations of 100 μM, 600

μM, and 2.0 mM are shown in figure 2.57, where a significant upfield shift of protons A-E of the naphthalene moieties is observed due to aggregation of **$BD_2$**.

**e. Fluorescence aggregation studies.** **$BD_2$** samples were prepared in 10 mM borate buffer, pH 9, at varying concentrations, and when applicable, a 3-fold excess of H-$KMe_3$-G-$NH_2$ was added. Both **$BD_2$** and KMe3 concentrations were determined by mass. Fluorescence scans were obtained at 25 °C in a 1 cm pathlength quartz microcuvette, using and excitation wavelength of either 240 nm or 278 nm and collecting an emission spectra from either 260-470 nm or 300-350 nm respectively.

## E. Selective recognition of asymmetric dimethylarginine and trimethyllysine: *receptor $A_2D$*

### i. Results and discussion

**a. Unbiased library templation studies.** Monomers **A**, **B**, and **D** can all be deemed "successful" building blocks, in that they have each been shown to contribute to the recognition of methylated PTMs, reacting with each other through disulfide exchange to construct the small molecule receptors *rac*-**$A_2B$**, *meso*-**$A_2B$**, and **$BD_2$**. On the contrary, monomers **C**, **E**, **G**, and **H** have been included in library screens, but have yet to be incorporated into a receptor that is significantly amplified in the presence of a methylated PTM. To determine whether any new receptors could be identified which incorporate all three of these "successful" building blocks in the generation of more potent, more selective PTM receptors, monomers **A**, **B**, and **D** were investigated in a single DCL.

While this **ABD** DCL has the potential to form and amplify novel PTM receptors, amplification of the previously identified receptors, *rac*-**$A_2B$**, *meso*-**$A_2B$**, and **$BD_2$**, should

also be expected to different extents depending on the PTM template. This could serve as a competition assay to some degree, yet it must be recognized that having multiple receptors amplified and therefore multiple equilibria at play in a single library complicates the system. As a result of this competition between receptors for specific building blocks, the efficiency of amplification in this case may not correlate directly with the strength of binding.[57] This breakdown in the correlation between amplification and binding efficiency occurs largely as a result of the tendency of DCLs to maximize the binding interactions of the entire library.[58] Amplification of a specific receptor may be less for one PTM than another, not necessarily because binding is weaker, but because a tighter binder has significantly depleted the supply of a required building block, for example. In addition, competition between macrocycles for specific building blocks can also result in a competitive advantage for those library members that are not dependent on those building blocks.[59]

The three building blocks were combined in equal amounts (2.5 mM each) at pH 8.5 (Figure 2.71). Initially, only the higher methylation state templates, KMe3, sRMe2, and aRMe2, were investigated, and thus three DCLs contained a PTM template at a concentration of 7.5 mM, while no template was added to the fourth DCL. The libraries were allowed to oxidize in capped HPLC vials at room temperature, and their compositions were analyzed by LC-MS at various time points.

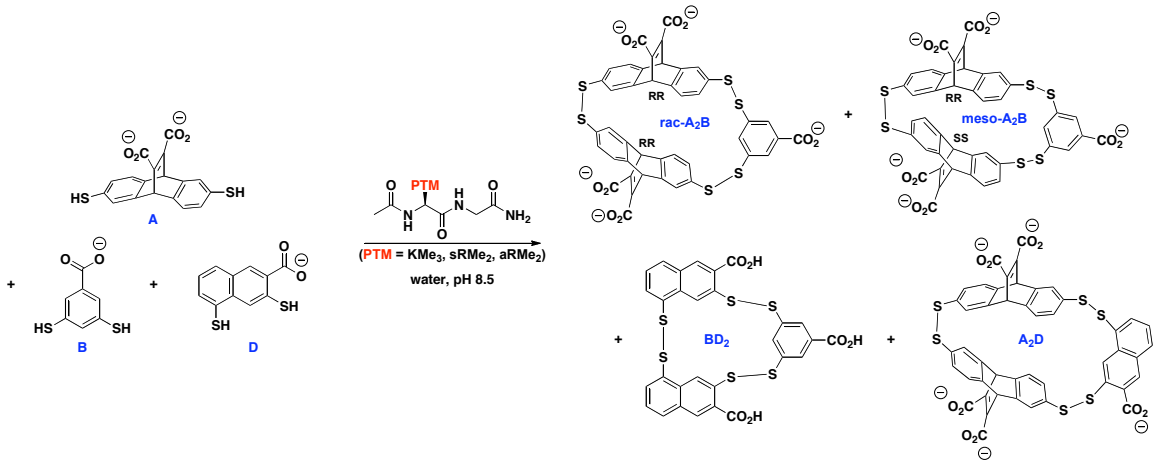



**Figure 2.71.** Unbiased DCL of dithiol monomers **A**, **B**, and **D** resulting in amplification of receptors *rac*-$\mathbf{A_2B}$, *meso*-$\mathbf{A_2B}$, $\mathbf{BD_2}$, and $\mathbf{A_2D}$ upon addition of a $KMe_3$ or $RMe_2$ dipeptide guests.

Initial library analysis revealed far more complex chromatograms relative to the previous library screens, with significantly more co-elution of library members. This highlights the limits associated with LC/MS as a detection method for amplification in dynamic combinatorial libraries, as libraries can very quickly become too complex for adequate resolution. Initially analyzed under conditions that were optimized for resolution of the library as a whole, in the absence of guest, numerous major constituents were detected, as well as many smaller peaks. Upon introducing a trimethyllysine dipeptide, the amplification of multiple peaks was observed, dominantly *rac*-$\mathbf{A_2B}$, *meso*-$\mathbf{A_2B}$, and $\mathbf{BD_2}$, as was to be expected (Figure 2.72). Receptor $\mathbf{BD_2}$ co-elutes with macrocycle $\mathbf{A_2B_3}$ in both templated and untemplated libraries, and therefore the exact extent of amplification is difficult to ascertain. In addition, the more subtle amplification of a fourth new receptor was observed, identified as $\mathbf{A_2D}$. Interestingly, the formation and amplification of only a single $\mathbf{A_2D}$ isomer was observed.

In comparing the amplification in this **ABD** DCL in the presence of KMe3 to that of aRMe2, as predicted, receptors *rac*-$\mathbf{A_2B}$, *meso*-$\mathbf{A_2B}$, and $\mathbf{BD_2}$ were all amplified to a lesser extent by aRMe2. In contrast, however, $\mathbf{A_2D}$ appeared to be amplified to a significantly greater extent when the library was templated with aRMe2 (Figure 2.72). This is the first time that a dimethylated Arg PTM was found to amplify a potential receptor to a greater extent than KMe3. By replacing the phenyl based monomer **B** with the naphthalene containing monomer **D** in $\mathbf{A_2B}$, it could be envisioned that the size of the binding pocket is increasing, in turn creating a tighter lock-and-key fit between $\mathbf{A_2D}$ and aRMe2 than was possible with $\mathbf{A_2B}$. While this amplification could indeed be the result of tighter binding to aRMe2, the degree of $\mathbf{A_2D}$ amplification in both cases may also be influenced by the extent of competition for building blocks **A** and **D** with the other amplified receptors.

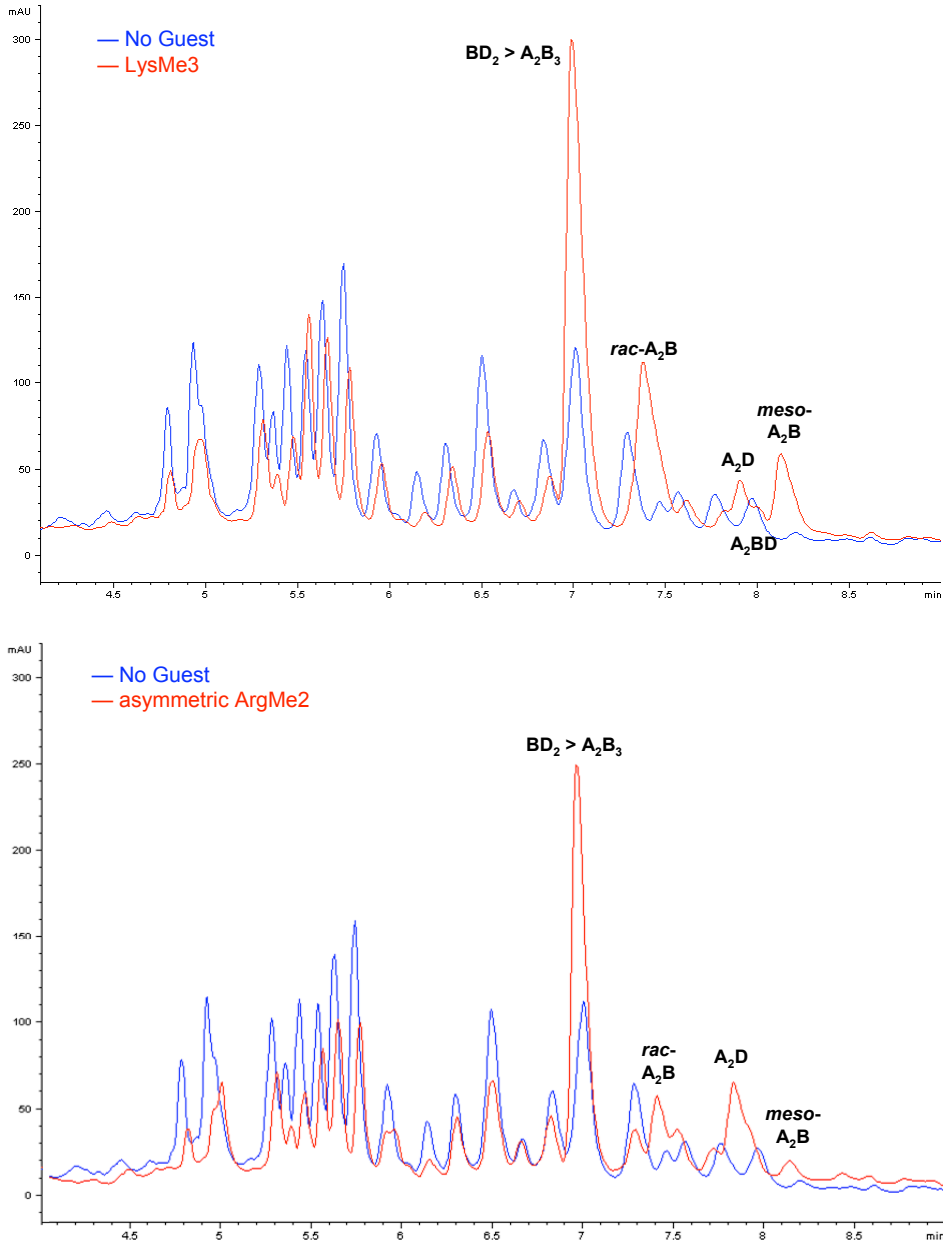



**Figure 2.72.** The analytical HPLC trace at 254 nm of a DCL consisting of monomers **A**, **B**, and **D** (2.5 mM each), untemplated (blue) and in the presence of KMe3 (top, red) and aRMe2 (bottom, red).

As there is some degree of co-elution of $\mathbf{A_2D}$ and $\mathbf{A_2BD}$, the LC/MS analysis conditions were optimized for baseline resolution of $\mathbf{A_2D}$ with little concern as to the resolution of the rest of the library. By doing so, a more accurate picture of the extent of amplification was obtained. While a close to undetectable concentration of $\mathbf{A_2D}$ was

present in the untemplated library, curtailing the integration accuracy and therefore the calculated percent change, a clear amplification trend was observed. Both dimethyl arginine PTMs amplified **$A_2D$** to a greater extent than LysMe3, with more favorable recognition of aRMe2 than sRMe2 (Figure 2.73).

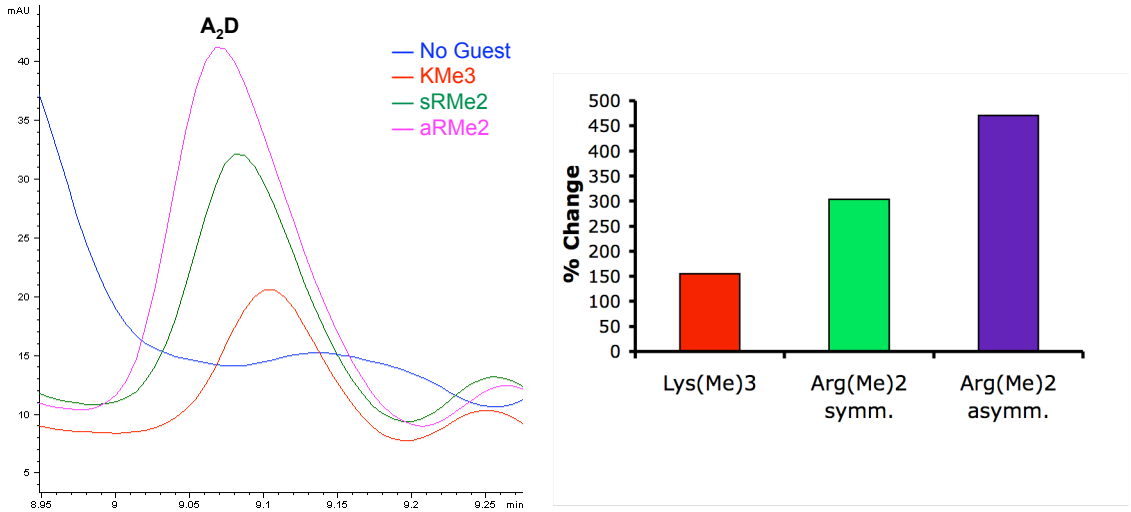



**Figure 2.73.** The extent of amplification of **$A_2D$** in an **ABD** library with PTM dipeptide guests (the **$A_2D$** HPLC peak at 254 nm on the left, and the calculated percent change on the right).

**b. Biased library templation studies.** While the **ABD** library results take into account multiple equilibria and are influenced by the competition between four receptors for certain building blocks, a biased library with only monomers **A** and **D** should allow for a more accurate evaluation of the recognition of methylated PTMs by **$A_2D$**. A DCL was set up biased towards the formation of **$A_2D$** in which building blocks **A** and **D** were mixed in a 2:1 ratio (7.5 mM total). At this point the selectivity of **$A_2D$** for the lower methylation states of lysine and arginine was also investigated, as the unmethylated lysine and arginine control peptides had not yet been evaluated. By biasing the library towards the formation of **$A_2D$**, this receptor is also formed to a larger extent, allowing for

a better visualization of the amplification as well as a more accurate integration of the peak areas.

A diverse library was generated in the absence of a template, with **$A_2D$** constituting only 3.3% of the total library composition. Upon templation with asymmetric RMe2, **$A_2D$** is strongly amplified and clearly becomes the largest species in the library (Figure 2.74). The amplification by symmetric RMe2 and RMe however is quite weak in comparison, showing a similar affinity of **$A_2D$** for both sRMe2 and RMe. This implies that **$A_2D$** exhibits quite impressive selectivity for aRMe2 over sRMe2. It is likely that only one methyl group of sRMe2 significantly contributes to binding, resulting in comparable amplification in the presence of both sRMe2 and RMe. It is possible that the pocket of **$A_2D$** is not large enough to accommodate sRMe2, whereas it fits much more tightly around aRMe2 where both methyl groups are in close proximity and can therefore make favorable contacts with the receptor.

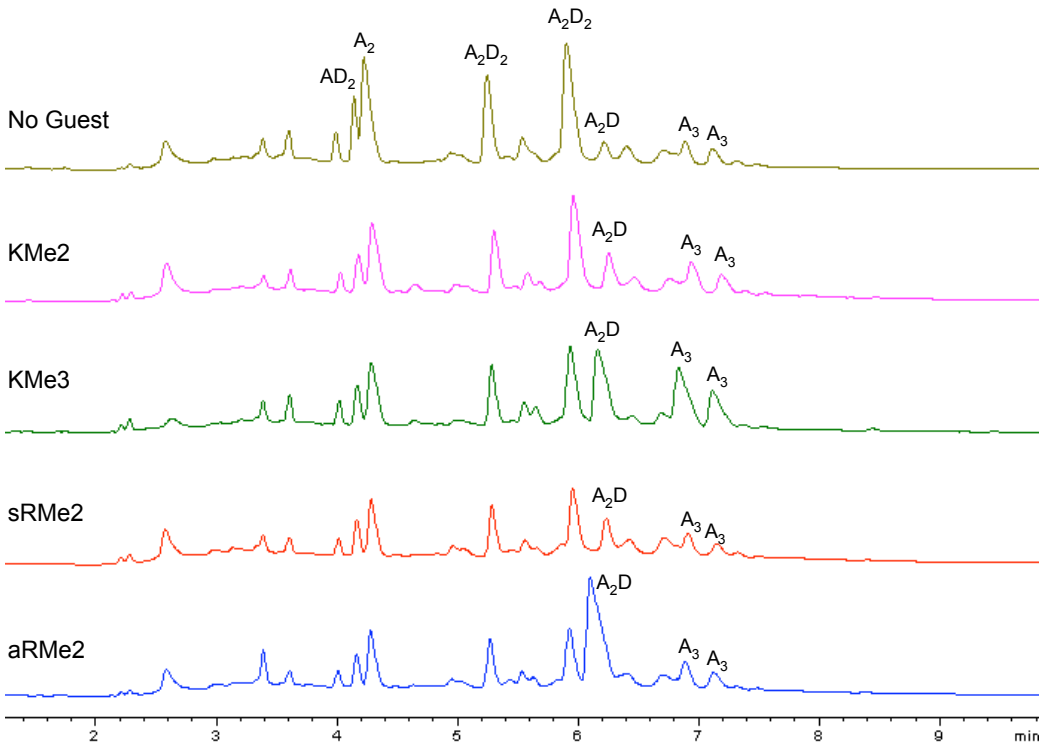



**Figure 2.74.** The analytical HPLC trace at 254 nm of a biased DCL consisting of monomers **A** (5 mM) and **D** (2.5 mM), untemplated (gold) and in the presence of KMe2 (magenta), KMe3 (green), sRMe2 (red), and aRMe2 (blue).

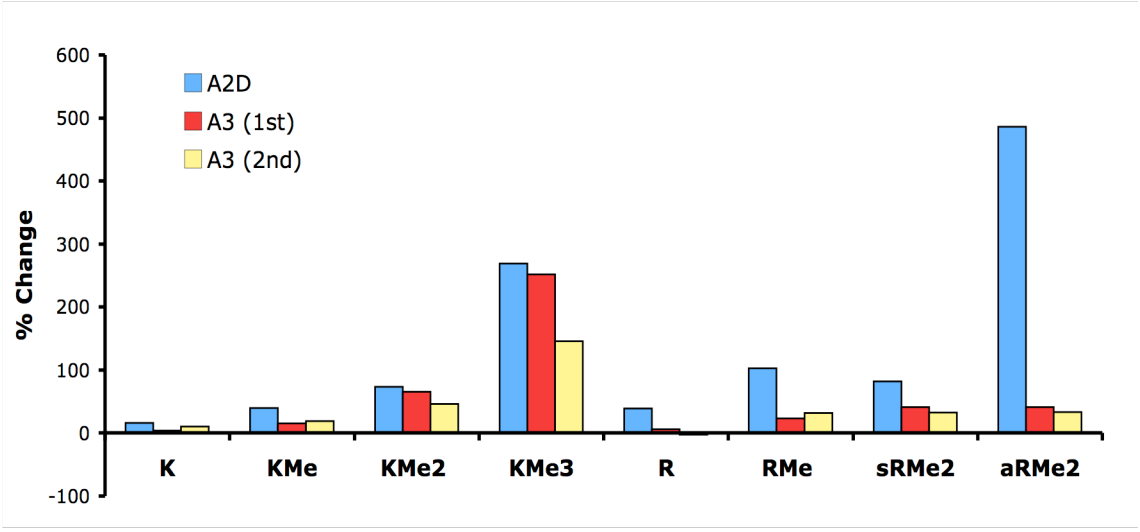



**Figure 2.75.** Amplification results of an **$A_2D$** biased library templated with all lysine and arginine dipeptide guests. The extent of amplification of **$A_2D$** (blue) and both **$A_3$** isomers (red and yellow) is shown.

In the presence of KMe3, **$A_2D$** is now amplified to a greater extent in the biased library relative to the unbiased **ABD** library. As a result of the high affinity of KMe3 for both **$A_2B$** and **$BD_2$**, it is likely that in the unbiased library both of these receptors were out-competing **$A_2D$** for the required monomers, resulting in a smaller observed amplification of **$A_2D$**. Asymmetric RMe2, however, binds **$A_2B$** and **$BD_2$** with very weak affinity, and therefore this competitive effect was minimal in comparison and a less drastic change in amplification between the unbiased and biased libraries was observed. The amplification of **$A_2D$** was dependant on the extent of lysine methylation, increasing with increasing methylation, suggesting a significant selectivity for KMe3 over both KMe2 and KMe (Figure 2.75). This increased affinity for KMe3 is likely due to its larger surface area, greater hydrophobicity, lack of desolvation penalty, and ability to form stronger cation-π interactions.

In addition to the amplification of **$A_2D$**, upon templation with KMe3 a similar degree of amplification of **$A_3$** was also observed in this library. Although to a much lesser magnitude, this is also observed with KMe2. Macrocycle **$A_3$** exists as two diastereomers, the major being the diastereomer in which the subunits have different chiralities, and the minor being the diastereomer in which all three subunits have the same chirality, as reported previously.[43] This again results in a competitive equilibrium in these lysine PTM templated libraries, as both receptors **$A_2D$** and **$A_3$** are in competition for monomer **A,** which may in turn breakdown the correlation between amplification and binding efficiency. As this is not the case for the aRMe2 containing DCL, it is again difficult to directly compare the amplification of **$A_2D$** by KMe3 versus aRMe2.

The observed amplification of $\mathbf{A_3}$ can also be used to make predictions about the affinity of $\mathbf{A_2D}$ for KMe3. In re-evaluating the $\mathbf{A_2B}$ biased libraries templated with KMe3, only a trivial amount of amplification of both $\mathbf{A_3}$ diastereomers is detected, indicating that both $\mathbf{A_2B}$ receptors are the more "fit" receptors. This is based on the idea that strong binders survive and are amplified while the weaker ones perish, and it is this principle that is considered to be the foundation of DCC.[57] The amplification of $\mathbf{A_2D}$ does not dominate over the amplification of the $\mathbf{A_3}$ isomers to the same degree, signifying that $\mathbf{A_2D}$ and $\mathbf{A_3}$ are both weaker KMe3 binders than *rac*-$\mathbf{A_2B}$ or *meso*-$\mathbf{A_2B}$. The extent to which $\mathbf{A_2D}$ binds more weakly than $\mathbf{A_2B}$, however, cannot be determined from these experiments.

**c. Preparative scale biased libraries and isolation of $\mathbf{A_2D}$.** To pursue further studies of $\mathbf{A_2D}$ regarding its affinity and selectivity for aRMe2, preparation and isolation of this macrocycle was required. Large scale biased libraries were prepared with a total volume of 10-20 mL and a ratio of **A** : **D** : guest of 2 : 1 : 3, with a total monomer concentration (**A** + **B**) of either 7.5 or 10 mM. Previously the affinity of $\mathbf{A_2D}$ for methylisoquinoline in comparison to aRMe2 was investigated on an analytical scale, and it was determined that methylisoquinoline was a superior template resulting in a higher concentration of $\mathbf{A_2D}$. Consequently, due to $\mathbf{A_2D}$'s apparent high affinity for methylisoquinoline and our ability to prepare it cheaply on a much larger scale than an aRMe2 peptide, these libraries were templated with methylisoquinoline. The three solids were combined with the appropriate amount of water, titrated to pH 8.5 with NaOH and HCl, and let stir in a capped scintillation vial for 1-3 weeks before being purified on a semi-preparative HPLC. The purity was confirmed in each case by LC/MS. Relative to

the synthesis and purification of *rac*- and *meso*-**A$_2$B**, the preparation of **A$_2$D** afforded higher yields, comparable to that of **BD$_2$**, still in low milligram quantities.

**d. Characterization of A$_2$D.** Since monomer **A** is used as a racemic mixture and monomer **D** lacks a plane of symmetry (unlike monomer **B**), there are two sets of possible **A$_2$D** diastereomers (Figure 2.76). In contrast to the amplification of both **A$_2$B** stereoisomers, the amplification of only a single **A$_2$D** diastereomer in the presence of methylated PTMs was observed, showing a surprisingly high degree of selectivity and indicating significant structural differences between the **A$_2$D** isomers. It can also be assumed that one isomer is significantly more thermodynamically stable than the others due to the fact that in the untemplated library a statistical ratio of the isomers is not observed. Instead, one peak corresponding to the mass of **A$_2$D** is significantly more dominant, and often the only **A$_2$D** peak in high enough concentration to be detected by MS. It has yet to be determined if the amplification of a single isomer is truly due to the fact that only one isomer is structurally capable of binding aRMe2 and therefore gets amplified, or if the other isomers are not present at all at thermal equilibrium or exist in such low concentrations that they go unnoticed.

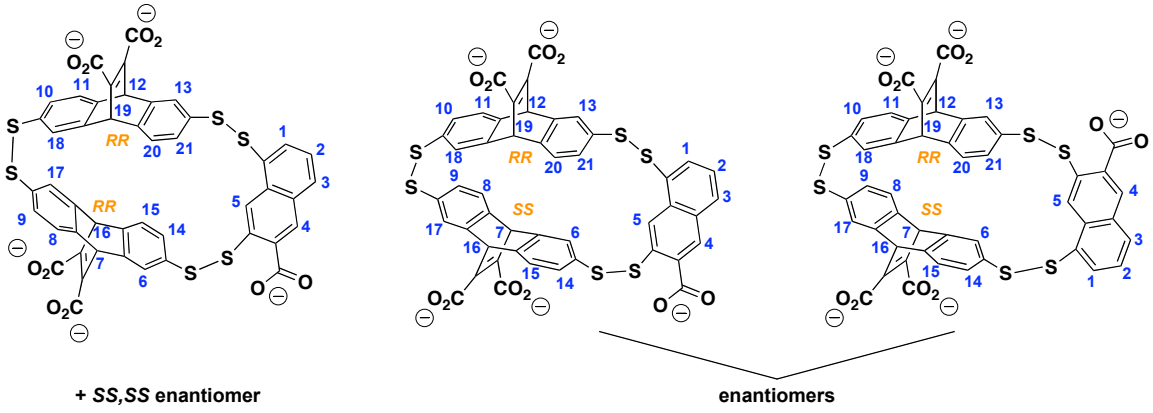



**Figure 2.76.** Isomers of **A$_2$D**.

The purified **$A_2D$** receptor was analyzed by NMR to gain further insight as to the macrocycle's purity and structure. Twenty-one distinct signals were expected for a single **$A_2D$** isomer, which is what was revealed in the 1D spectra, confirming that only a single isomer was present in solution. Low temperature experiments proved crucial to achieve sufficient resolution to distinguish all 21 protons. Furthermore, changes in the concentration of **$A_2D$** did not result in any noticeable changes in the proton chemical shifts, eliminating the possibility of **$A_2D$** aggregation as a potential problem for future studies.

In combination with the 1D spectra, NOESY, TOSCY, and COSY experiments revealed a great deal of information about the conformation of the **$A_2D$** receptor. Most notable were the chemical shifts of protons 2 and 3 of monomer **D**, shifted up to approximately 3.1 ppm and 2.9 ppm respectively (Figure 2.77). This is almost 5 ppm further upfield from where these proton signals appear in the spectra of the monomer in isolation, indicating a significant interaction with the remainder of the macrocycle. It is proposed that this portion of the naphthalene moiety is pointing towards the inside of the receptor cavity, sandwiched between both **A** molecules and packing into the hydrophobic pocket. Although to a lesser extent, protons 1 and 4 also appear upfield from where they would be expected, again indicating that monomer **D** is packing up against the hydrophobic portions of both **A** monomers. This was confirmed by the presence of NOESY cross peaks between both protons 1 and 4 of **D** and protons 6 and 13 (Figure 2.78, green arrows).

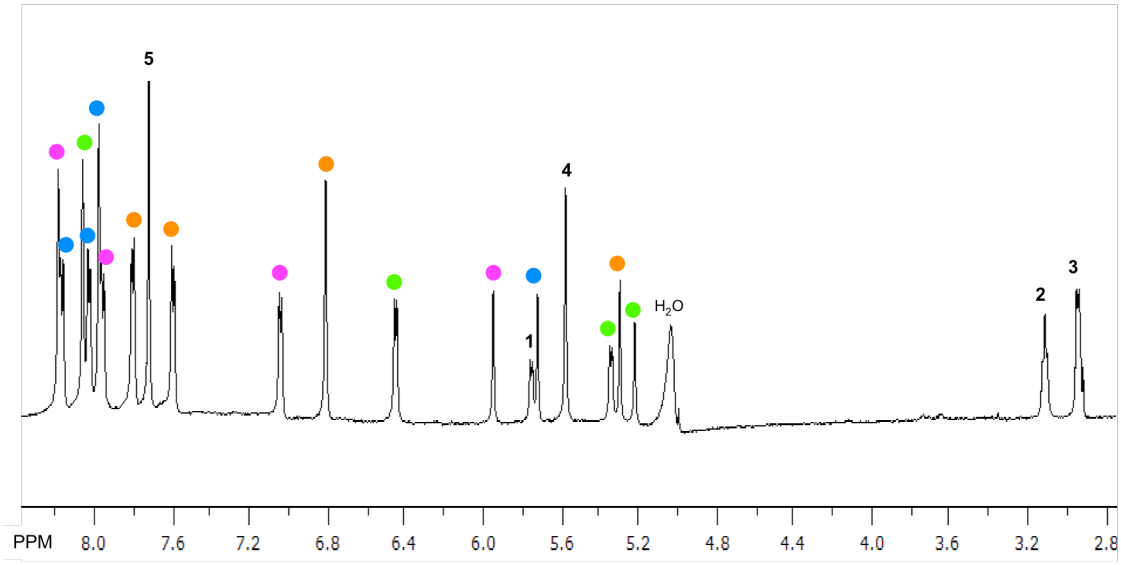



**Figure 2.77.** $^{1}$H NMR of **$A_2$D** at 5 °C in borate buffer. The numbered protons correspond to those assigned in monomer **D**. The blue and orange protons are each from an **A** molecule on the same face of **$A_2$D**, and are in close proximity (protons 14 – 21, figure 2.78), whereas the green and magenta protons are on the other face of **$A_2$D** and are further from each other (protons 6 – 13).

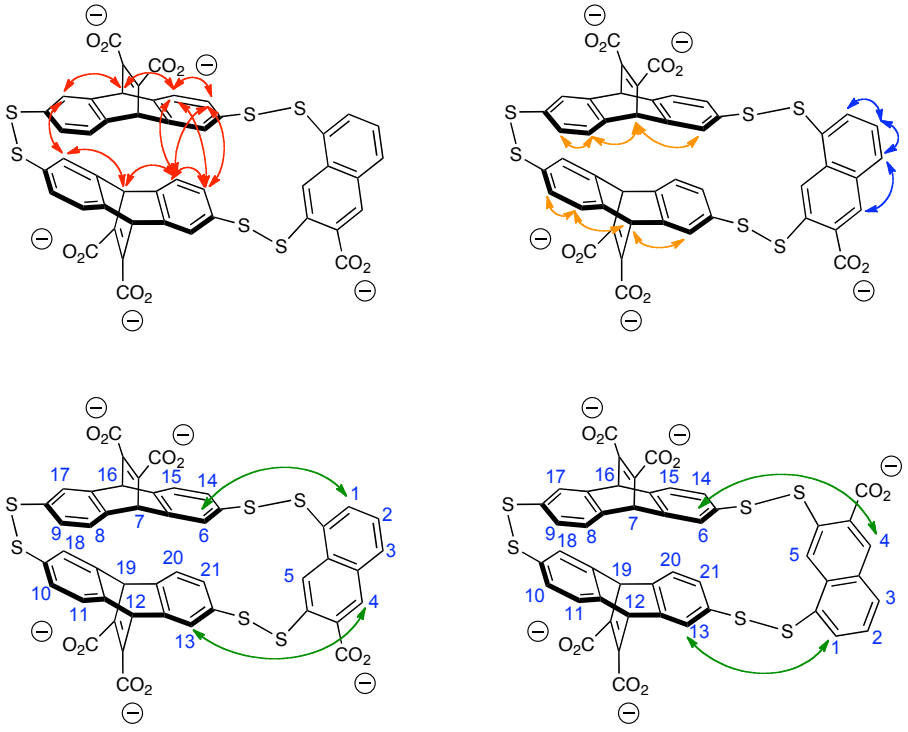


**Figure 2.78. $A_2$D** NOESY cross peaks within each monomer unit and between monomers. Cross peaks are observed between protons 1 and 6 and 4 and 13 in one enantiomer (of the *RR,SS*/*SS,RR* pair), and between 1 and 13 and 4 and 6 in the other.

Furthermore, **$A_2D$** is predicted to be the *RR,SS*/*SS,RR* enantiomers (Figure 2.76). This is due to the presence of cross peaks between four monomer **A** aromatic doublets all in close proximity on one face of the molecule (for example 14, 15, 20, and 21), as indicated by the NOESY spectra (Figure 2.78, red arrows). A cross peak is also present between protons 17 and 18 further confirming this hypothesis. While both **A** monomers appear to be in close proximity on one face of the molecule, as one can "walk" in a complete circle from proton 14 to proton 21 in the NOESY spectra, the other sides of both **A** molecules appear to be further apart from each other (protons 6-14), creating somewhat of a binding cleft. These NMR experiments provide substantial evidence that **$A_2D$** is forming a good binding cavity for methylated PTMs, however it is not known to what degree **$A_2D$** maintains this structure upon binding.

**e. NMR analysis of asymmetric dimethyl arginine binding to $A_2D$.** NMR was also used to investigate the recognition of asymmetric ArgMe2 by **$A_2D$**. The acylated Ac-aR$Me_2$-G-$NH_2$ dipeptide was first characterized at 5 °C in aqueous buffer. Subsequently, the peptide was analyzed under the same conditions in the presence of an excess of **$A_2D$** to ensure that all of the aRMe2 peptide was bound. In addition to noticeable broadening of both the **$A_2D$** and peptide signals upon binding, a clear shift of the dipeptide protons was observed, confirming that the methylated arginine is indeed bound to the hydrophobic cleft of **$A_2D$**. While the largest upfield shift was observed for the two methyl groups which are indistinguishable by NMR, the extent of shifting decreases moving down the Arg side chain towards the peptide backbone (Figure 2.79). The methyl groups are upfield shifted 0.914 ppm, which is quite comparable to the shifts observed in the recognition of KMe3 by **$A_2B$** and **$BD_2$**.

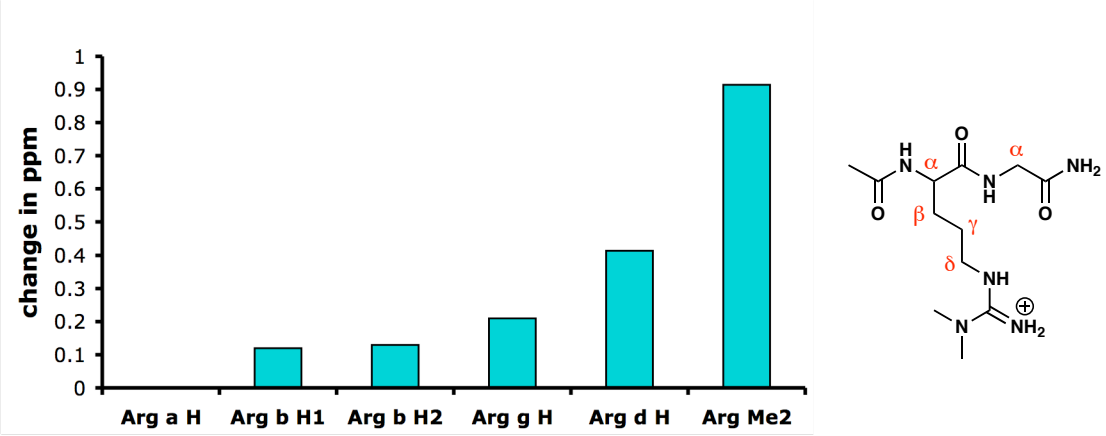



**Figure 2.79.** Upfield side-chain chemical shifts of Ac-aRMe$_2$-G-NH$_2$ upon binding to **A$_2$D** at 5 °C. Conditions: 50 mM borate pD 9.0, referenced to DSS.

For comparison, the extent of shifting of the KMe3 methyl groups upon binding **A$_2$D** was assessed. If both aRMe2 and KMe3 are binding in a comparable fashion, it would be expected that the methyl groups of both PTMs would be affected similarly. At 5 °C, the three methyl groups on lysine were found to shift 0.985 ppm, which is very similar to the observed shift of the arginine methyl groups. These results confirm that the key binding interaction in both cases is between the methylammonium group and the aromatic pocket of **A$_2$D**.

**f. Binding studies by fluorescence quenching.** In order to evaluate the affinity of **A$_2$D** for the different methylated PTMs, as well as the extent of selectivity, for example between aRMe2 and sRMe2 or between aRMe2 and KMe3, binding studies were conducted. The determination of binding constants for **A$_2$D** will allow for an interesting comparison between **A$_2$D** and **A$_2$B**, and reveal more information regarding the effect of replacing the phenyl ring in **A$_2$B** with the naphthalene moiety in **A$_2$D**. Binding studies were initially attempted via fluorescence anisotropy to allow for a direct comparison to the affinities determined for **A$_2$B**. The histone tail peptides previously synthesized

containing residues 5-12 of the histone 3 tail and a methylated residue at either Arg8 or Lys9, as well as an N-terminal carboxyfluorescein were used to measure binding. The binding studies were performed under similar conditions as those for **$A_2B$**.

In addition to measuring the fluorescence anisotropy for each experiment, a simple fluorescence intensity reading was performed on each sample containing both fluorescent peptide and a known concentration of **$A_2D$**. As the amount of peptide remains constant in each, this fluorescence reading is expected to remain unchanged throughout the experiment, despite the increasing concentration of **$A_2D$**. This indicates that **$A_2D$** is not interacting directly with the fluorophore to cause any form of undesired fluorescence quenching, and therefore the anisotropy readings are based solely on the variations in parallel and perpendicular fluorescence due to changes in fluorescein rotation upon binding.

In initial experiments with different PTM peptides in which the level of fluorescein fluorescence is evaluated throughout the experiment, a change in fluorescence was observed. This change was generally a drop in fluorescence intensity with increasing concentrations of **$A_2D$**, indicating some degree of fluorescein quenching by **$A_2D$**. While naphthalenes have been shown to quench the fluorescence of other naphthalene molecules in the case of **$BD_2$**, it should not be entirely surprising that monomer **D** would also serve as a quencher of other fluorophores such as fluorescein. Due to the fact that the extent of fluorescence quenching varied from one PTM peptide to another, this revealed that this was not a systematic interaction. As a result, the anisotropy binding data was being affected dissimilarly for the different PTM peptides, and this no longer

allowed for an accurate evaluation and comparison of the binding affinities via fluorescence anisotropy.

Interestingly, the degree of fluorescein quenching was greatest for the H3 peptides containing either aRMe2 or KMe3, whereas the peptides containing lower methylation states exhibit little to no fluorescence quenching. Based on prior studies, both aRMe2 and KMe3 are expected to bind **$A_2D$** most tightly, and this implies that the extent of quenching is related to the PTM binding affinity. Most likely, the binding event occurs first which puts the naphthalene in close proximity to the fluorescein, and the fluorescence is quenched as a result. If the quenching were not binding induced, the same degree of quenching would be expected for all PTM peptides, regardless of their affinity for **$A_2D$**.

Since the degree of quenching is a reflection of binding affinity, this allowed for binding to be quantified by fluorescein quenching. The fluorescence intensity of an H3 control peptide with both Arg8 and Lys9 mutated to Gly was first measured in a similar fashion in the presence of increasing concentrations of **$A_2D$**. In each case the fluorescence of this control peptide, which does not bind and therefore exhibits no quenching, was subtracted from that of each PTM peptide fluorescence to ensure that the fluorescence change analyzed was that solely due to **$A_2D$** binding to the PTM of interest. The change in fluorescein fluorescence for peptides containing all methylation states at either Arg8 or Lys9 was evaluated.

The H3 aRMe2 peptide was found to bind **$A_2D$** with a binding affinity of 38 μM, which is slightly weaker than the affinity determined for the binding of **$A_2B$** with KMe3, yet is still comparable to the affinity of many natural PTM protein receptors. Equally

significant is the observed selectivity, as **$A_2D$** was found to bind the H3 aRMe2 peptide with greater than 6-fold selectivity over its symmetric RMe2 counterpart ($K_d$ = 236 μM). The ability of **$A_2D$** to discriminate between these closely related PTMs is quite promising, enhancing the potential usefulness of this small molecule receptor. It is clear that either the difference in size of these two RMe2 PTMs, or the localization of both methyl groups on a single nitrogen in aRMe2 play a significant role in enabling **$A_2D$** to differentiate between the two modifications. The H3 RMe peptide exhibited even weaker affinity, while no appreciable binding was observed for the unmethylated histone tail.

**Table 2.10.** Dissociation constants for H3 tail peptides with varying methylation states at R8 and K9 as determined by fluorescence quenching.[a]

| Peptide | **$A_2D$** $K_d$ (μM)[b] |
|---|---|
| **H3 asymmetric R8Me$_2$** | 38 ± 8 |
| **H3 symmetric R8Me$_2$** | 236 ± 39 |
| **H3 R8Me** | 474 ± 79 |
| **H3 K9Me$_3$** | 40 ± 7 |
| **H3 K9Me$_2$** | 95 ± 11 |
| **H3 K9Me** | 663 ± 234 |
| **H3 R8K9** | No Binding |

[a] Conditions: 27 °C, 10 mM phosphate buffer, pH 8.5. [b] Errors are from the fit.

Interestingly, the binding affinity of **$A_2D$** for KMe3 was found to be quite comparable to that of aRMe2, showing little distinction between the two quite different modifications. Despite their similarity in binding affinity, it is likely that various favorable interactions are contributing to a different degree in each case, however producing a similar end

result. For example, KMe3 likely results in a stronger cation-π interaction with the receptor, whereas aRMe2 may result in a tighter lock-and-key type fit with **$A_2D$**. This comparable affinity would not have otherwise been realized based on the **$A_2D$** amplification data alone, which is likely a result of the competitive amplification of **$A_3$** in the case of KMe3, without which the amplification of both aRMe2 and KMe3 may have been comparable. This highlights the possible breakdown of the correlation between amplification and binding efficiency as a result of complex competing equilibrium in a single DCL, and the tendency of DCLs to maximize the binding interactions of the entire library. Similarly, **$A_2D$** was found to bind KMe2 with a dissociation constant of 95 μM, which is again higher than may have been predicted based on amplification data alone due to the competitive binding with **$A_3$**. The affinity continues to drop with decreasing lysine methylation, mimicking the binding trend of both **$A_2B$** and native protein receptors.

### ii. Conclusions

In addition to the identification of receptors for the higher methylation states of lysine, we have now identified a small molecule receptor exhibiting equally promising affinity for asymmetric RMe2. Although little is known regarding the affinity of aRMe2 protein binding domains, it can be predicted that the affinity of **$A_2D$** will not only parallel that of protein receptors, but also allow for its use in a variety of biological applications. As methylated arginine PTMs and their corresponding protein receptors remain less well understood, such small molecule receptors are especially attractive in both the

investigation of known sites of arginine methylation as well as the identification of new sites of methylation.

In comparing **$A_2D$** to both *rac*- and *meso*-**$A_2B$**, it is clear that replacing one monomer makes a substantial impact on the receptor's recognition capabilities. **$A_2D$** not only recognizes aRMe2 with about 16-fold greater affinity than *rac*-**$A_2B$**, but its affinity for all methylated lysine PTMs is slightly diminished due to the change in receptor structure, while its affinity for all methylated arginine PTMs is slightly enhanced. Although these small molecule receptors do not display an absolute selectivity for a single PTM, they exhibit a degree of selectivity that has not been demonstrated previously with small molecule PTM receptors. Furthermore, they display a level of affinity and selectivity that would be challenging to predict and achieve *a priori*, demonstrating the overall utility of DCC as a high-throughput method for developing such PTM receptors for biological applications.

### iii. Experimental Section

**a. Dynamic combinatorial library screens.** Overall, library screens were conducted similarly to those described previously. The relevant building blocks were individually dissolved in water with sufficient 1.0 M aqueous NaOH to bring them into solution, and the pH of each solution was then adjusted to 8.5. In the unbiased **ABD** libraries, aliquots of each monomer solution were combined in a 2 mL LC-MS vial to reach a final concentration of 2.5 mM of each monomer (7.5 mM total). In the biased **$A_2D$** libraries, aliquots of each monomer solution were combined in a 2 mL LC-MS vial to reach a final concentration of 5 mM **A** and 2.5 mM of **D**. When necessary, aliquots of

peptide guests dissolved in water were added to the reactions to reach a final concentration of 7.5 mM peptide. The vials were capped and analyzed at various time points.

**b. Analytical LC/MS.** LC-MS was carried out on an Agilent Rapid Resolution LC-MSD system equipped with an online degasser, binary pump, autosampler, heated column compartment, and diode array detector. All separations were performed using 5 mM $NH_4OAc$ $H_2O$-acetonitrile gradients at pH 5 and a Halo C18 column (4.6 × 100 mm, 2.7 micron). The MS was performed using a single quadrupole mass spectrometer via negative electrospray ionization. The reactions were monitored weekly (3 μL injections) until equilibrium was reached. The chromatography of both **ABD** and **$A_2$D** DCLs were carried out with gradient G (Table 2.11), using a gradient temperature, going from 40°C to 50°C, left to right. The peak areas were integrated at 254 nm and the percent change in concentration and/or amplification factors were calculated.

**Table 2.11.** Analytical LC methods use to analyze DCC libraries **ABD** and **$A_2$D** (Method G).

| **Gradient G** | | |
|---|---|---|
| **Time (min)** | **%B** | **Flow Rate (mL/min)** |
| 0.00 | 3.0 | 1.0 |
| 3.00 | 30.0 | 1.0 |
| 8.00 | 33.0 | 1.0 |
| 10.30 | 34.0 | 1.25 |
| 12.00 | 50.0 | 1.25 |
| 13.00 | 100.0 | 1.25 |
| 15.00 | 100.0 | 1.0 |
| 15.10 | 3.0 | 1.0 |
| 20.00 | 3.0 | 1.0 |

**c. Synthesis and isolation of $A_2$D.** Biased libraries were prepared on a 15 mL scale (**A**: 35.6 mg, 0.1 mmol, 6.67 mM; **D**: 11.8 mg, 0.05 mmol, 3.33 mM) templated

with methylisoquinoline triflate (44.0 mg, 0.15 mmol, 10 mM). Upon equilibration the libraries were neutralized and the receptors were isolated by semi-preparative HPLC. Approximately 0.5 mL injections were chromatographed using standard peptide synthesis mobile phases A and B (0-50% B 0-5 min, then 50-75% B 5-20 min) with a flow rate of 4.0 mL/min. A sharper **$A_2D$** peak was achieved with a column heater set to 40°C, but this was not necessary for separation from the library. The **$A_2D$** peak at 13.5 minutes was collected (Figure 2.80) and analyzed for purity by analytical LC-MS (Figure 2.81). Purified **$A_2D$** was lyophilized to powder and stored under nitrogen.

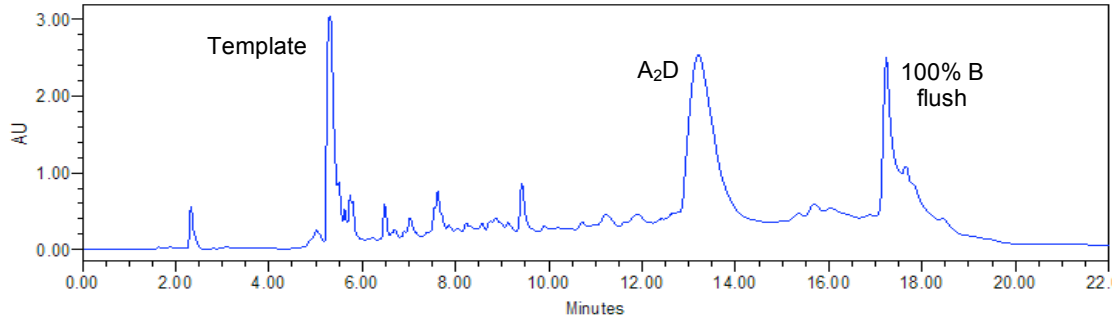



**Figure 2.80.** Semi-preparative HPLC trace of an **$A_2D$** biased library at 254 nm.

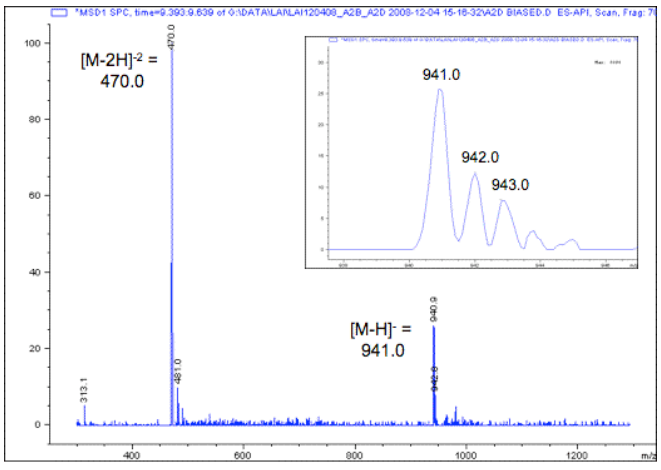



**Figure 2.81.** Mass spectra of purified **$A_2D$** (-ESI).

**d. NMR spectroscopy.** All 1D and 2D NMR samples were prepared and analyzed on a Varian Inova 600-MHz instrument similarly to those described previously, unless noted otherwise. For characterization, a 600 μM **$A_2D$** sample was prepared in 50 mM borate buffer at pD 9.25 containing 0.05 mM DSS (Figure 2.77). The 1D spectrum of **$A_2D$** was collected at 5 °C using 32K data points and 900 scans with a 3 second presaturation. $^1$H NMR (600 MHz): δ = 8.20 (s, 1H), 8.18 (s, 1H), 8.07 (s, 1H), 8.04 (d, 1H), 7.99 (s, 1H), 7.97 (d, 1H), 7.81 (d, 1H), 7.73 (s, 1H), 7.61 (d, 1H), 7.05 (d, 1H), 6.82 (s, 1H), 6.46 (d, 1H), 5.95 (s, 1H), 5.75 (d, 1H), 5.72 (s, 1H), 5.58 (s, 1H), 5.34 (d, 1H), 5.30 (s, 1H), 5.22 (s, 1H), 3.10 (dd, 1H), 2.93 (d, 1H). 2D NMR experiments used for structural analysis included COSY, TOSCY, and NOESY experiments.

To examine asymmetric RMe2 binding to **$A_2D$**, two samples, one with Ac-aR$Me_2$-G-$NH_2$ (0.65 mM) only, and the other Ac-aR$Me_2$-G-$NH_2$ (0.65 mM) bound to **$A_2D$** (1.0 mM), were prepared in 50 mM borate buffer pD 9.25. 1D spectra were collected for each 5 °C with 32K data points and 64 scans. All arginine protons were assigned using standard methods. An identical experiment was performed with Ac-K$Me_3$-G-$NH_2$ (0.65 mM) in the absence and presence of **$A_2D$** (1.0 mM).

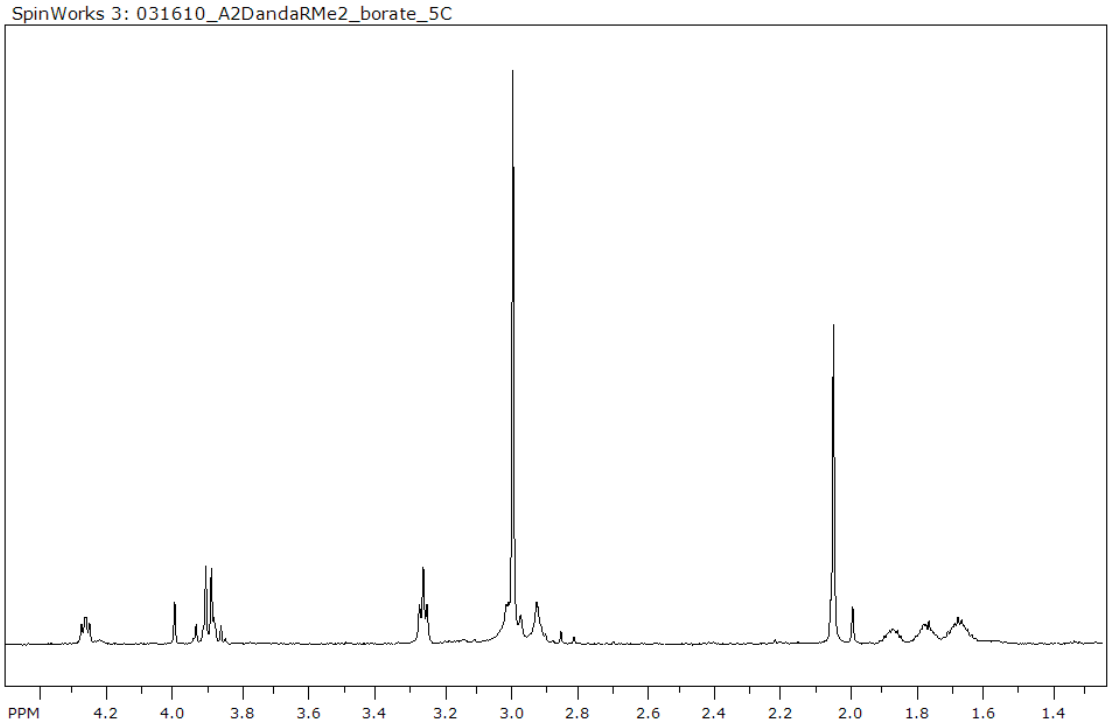



**Figure 2.82.** $^{1}$H NMR of Ac-aRMe$_2$-G-NH$_2$ at 5 °C.

**Table 2.12.** Proton chemical shift assignments for $^{1}$H NMR of Ac-aRMe$_2$-G-NH$_2$ at 5 °C.

| Residue | α | β | γ | δ | $Me_2$ |
|---|---|---|---|---|---|
| **aRMe$_2$** | 4.264 | 1.869, 1.768 | 1.674 | 3.263 | 2.996 |
| **G** | 3.901 | | | | |

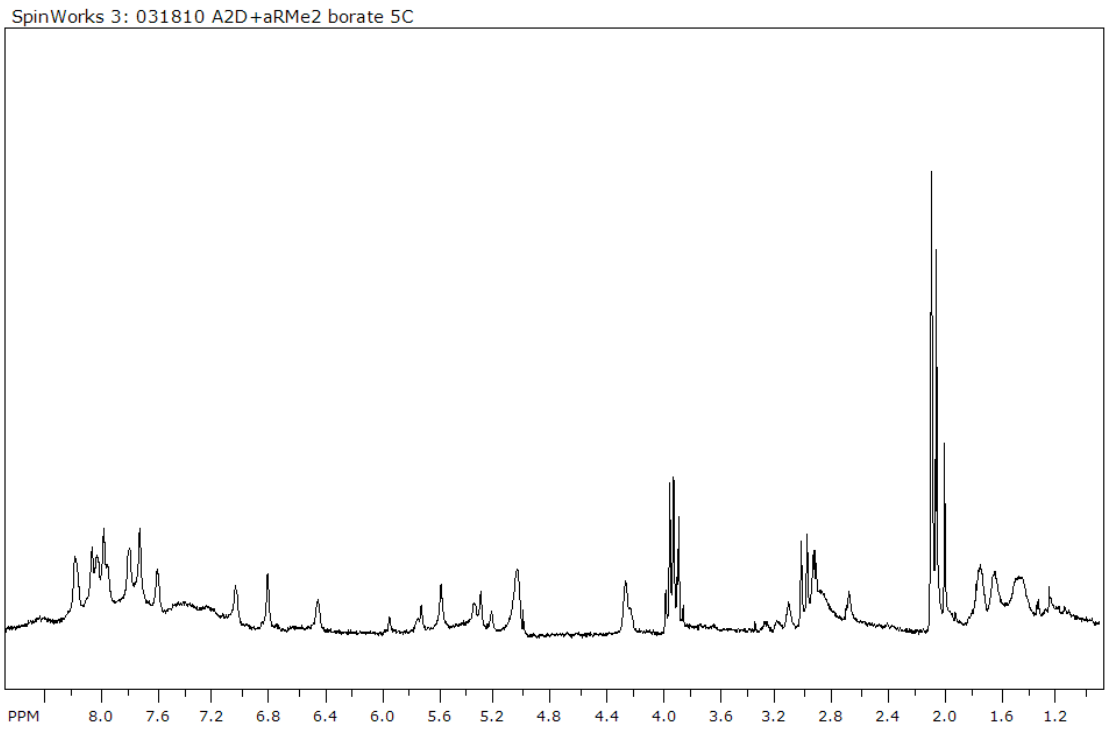



**Figure 2.83.** $^{1}$H NMR of H-K$Me_3$-G-$NH_2$ with **$A_2$D** at 5 °C.

**Table 2.13.** Proton chemical shift assignments for $^{1}$H NMR of Ac-a$RMe_2$-G-$NH_2$ bound to **$A_2$D** at 5 °C.

| Residue | α | β | γ | δ | $Me_2$ |
|---|---|---|---|---|---|
| **a$RMe_2$** | 4.266 | 1.746, 1.629 | 1.440 | 2.850 | 2.082 |
| **G** | 3.936 | | | | |

**e. Fluorescence quenching binding experiments.** Binding assays were performed with purified **$A_2$D** and fluorescein labeled histone 3 peptides. Peptides were dissolved in buffer (10 mM phosphate pH 8.5), and 25 μM solutions were prepared as determined by UV-Vis at 492 nm (ε = 78,000 $M^{-1}cm^{-1}$). A 0.25 mM stock solution of **$A_2$D** was prepared, with the solution concentration determined by mass. Assays were prepared in 96-well half-area plates with a total volume of 50 μL per well, containing 5 μM labeled peptide and increasing concentrations of **$A_2$D** (from 0 μM to 195 μM) in

buffer. Plates were spun down and allowed to incubate for at least 30 minutes before analysis. Fluorescence scans were measured on a PHERAstar plate reader (BMG Labtech) at approximately 27°C, using an excitation wavelength of 485 nm and an emission wavelength of 520 nm (Figure 2.84).

In each case the fluorescence scans for the control peptide, containing glycine at Arg8 and Lys9, were subtracted from the raw quenching data for each PTM experiment before fitting the data. Fluorescence titrations were plotted as a function of **$A_2D$** concentration and fit to the following equation for 1:1 binding in KaleidaGraph, where I is the observed fluorescence intensity, $I_o$ is the initial fluorescence intensity of the peptide, $I_\infty$ is the fluorescence intensity at binding saturation, [L] is the concentration of the added receptor, and $K_d$ is the dissociation constant. All measurements were taken in duplicate or triplicate.

$$I = (I_o + I_\infty([L]/K_d)) / (1 + ([L]/K_d))$$

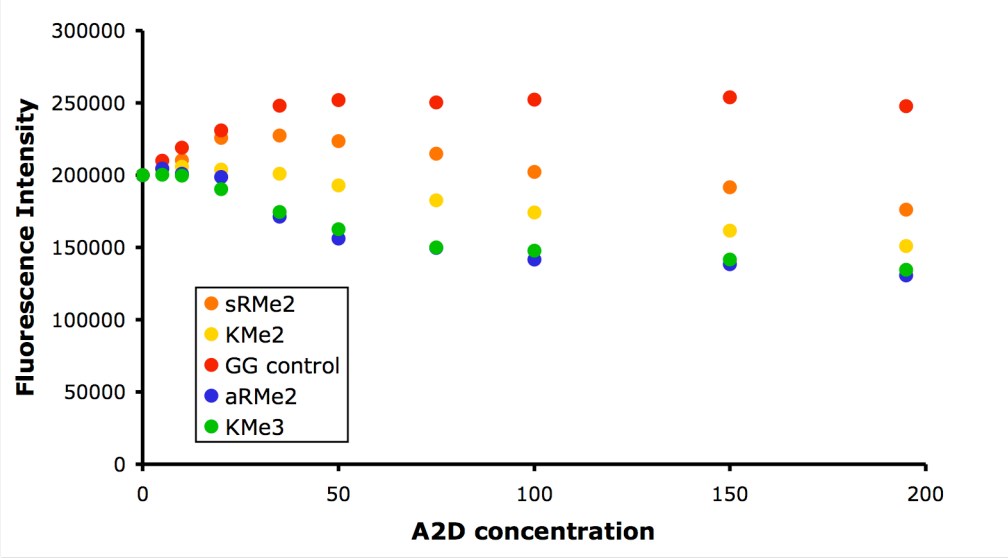



**Figure 2.84.** Representative uncorrected fluorescence intensity data prior to subtraction of the Gly8-Gly9 control peptide (Ac-QTA**GG**STG-$NH_2$). Fluorescence titration of G8-G9 (red), sRMe2 (orange), KMe2 (yellow), KMe3 (green), and aRMe2 (blue) with **$A_2D$**.

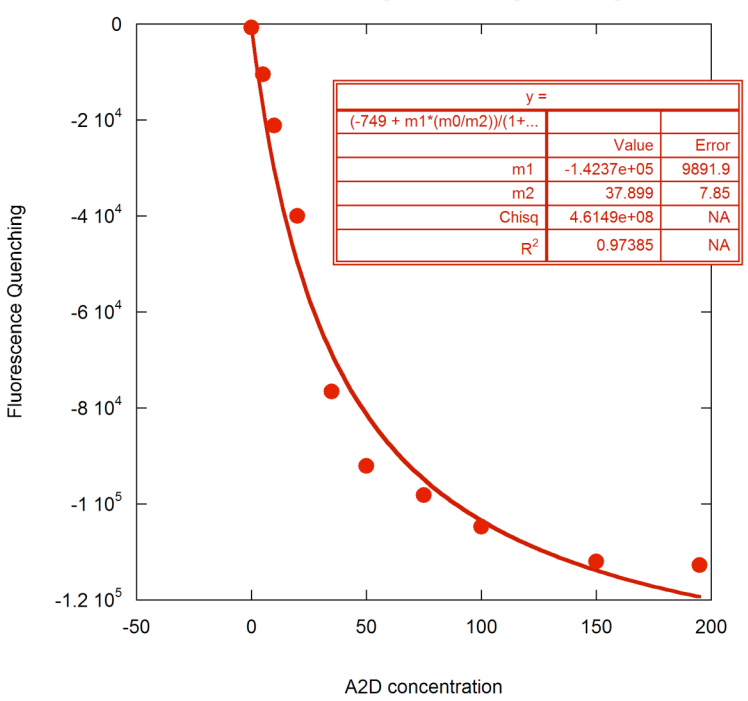



**Figure 2.85.** Fluorescence titration of H3 aRMe2 with **$A_2D$**. The data is corrected by subtracting the fluorescence of the H3 Ac-QTA**GG**STG-$NH_2$ control peptide.

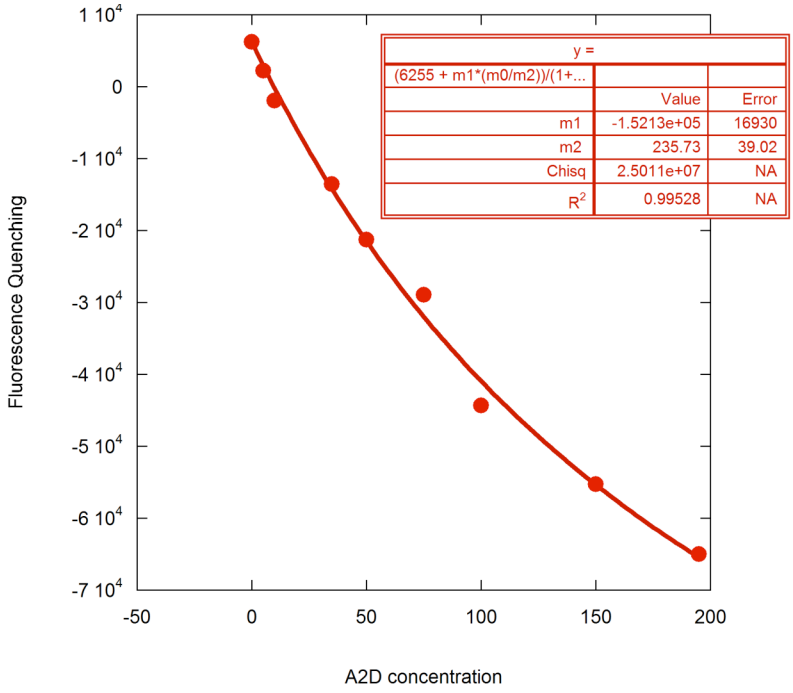



**Figure 2.86.** Fluorescence titration of H3 sRMe2 with **$A_2D$**. The data is corrected by subtracting the fluorescence of the H3 Ac-QTA**GG**STG-$NH_2$ control peptide.

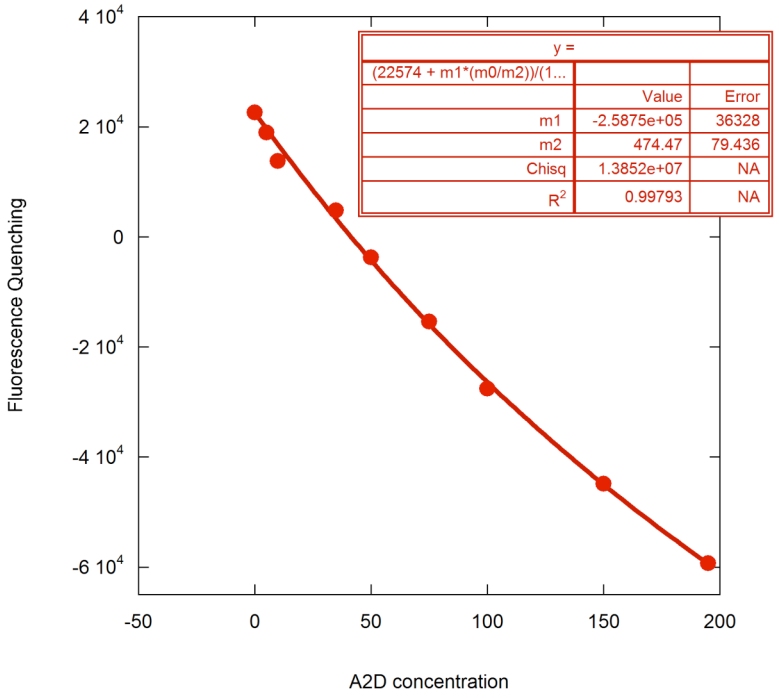



**Figure 2.87.** Fluorescence titration of H3 RMe with **$A_2D$**. The data is corrected by subtracting the fluorescence of the H3 Ac-QTA**GG**STG-$NH_2$ control peptide.

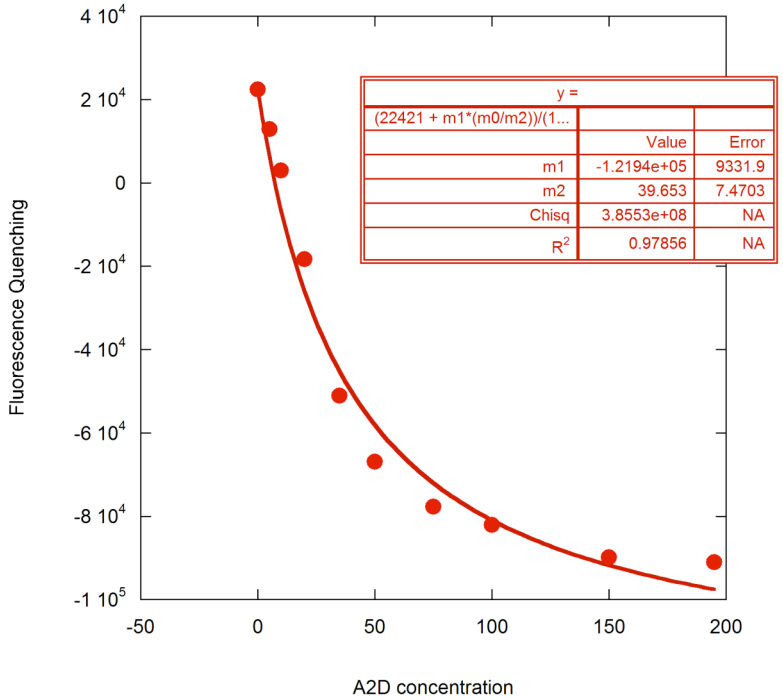



**Figure 2.88.** Fluorescence titration of H3 KMe3 with **$A_2D$**. The data is corrected by subtracting the fluorescence of the H3 Ac-QTA**GG**STG-$NH_2$ control peptide.

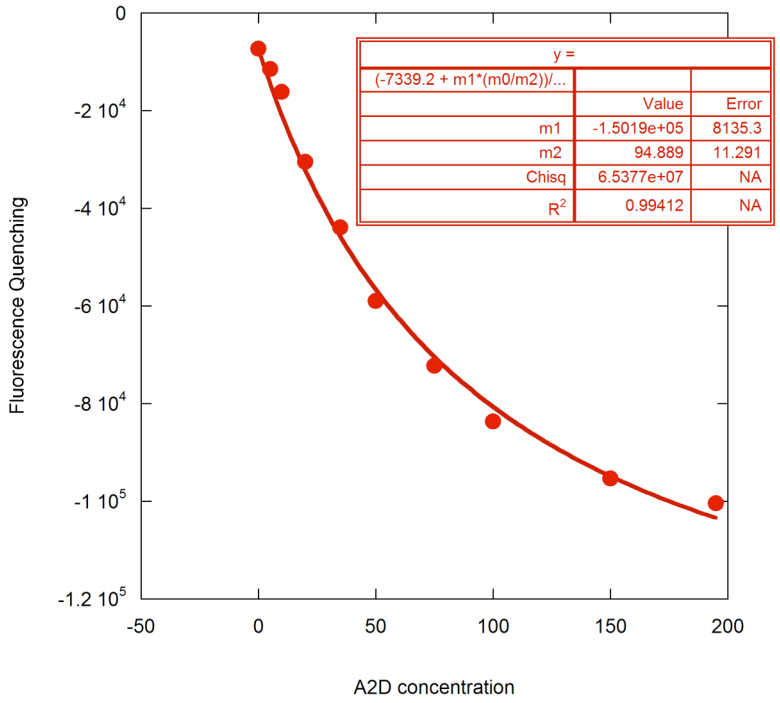



**Figure 2.89.** Fluorescence titration of H3 KMe2 with **$A_2D$**. The data is corrected by subtracting the fluorescence of the H3 Ac-QTA**GG**STG-$NH_2$ control peptide.

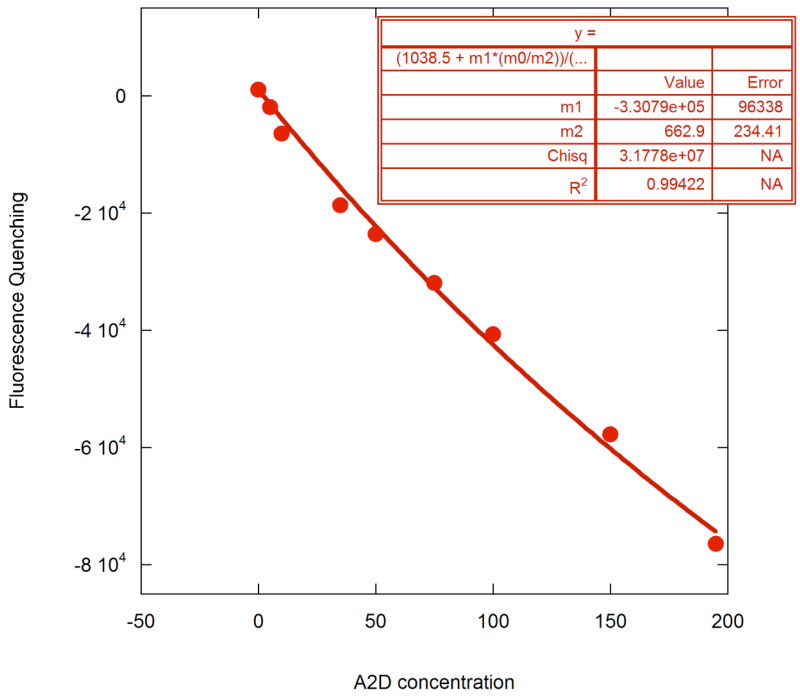



**Figure 2.90.** Fluorescence titration of H3 KMe with **$A_2D$**. The data is corrected by subtracting the fluorescence of the H3 Ac-QTA**GG**STG-$NH_2$ control peptide.

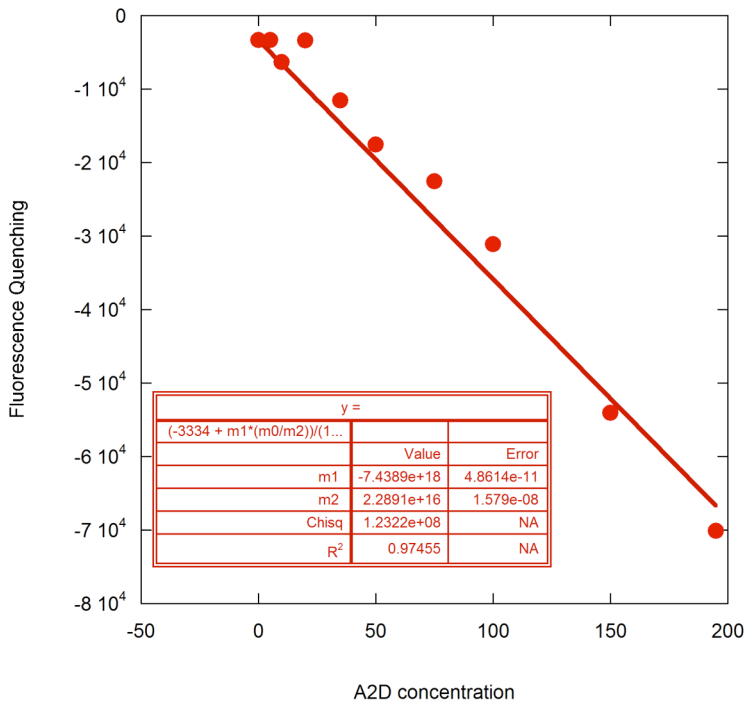



**Figure 2.91.** Fluorescence titration of H3 RK with **$A_2$D**. The data is corrected by subtracting the fluorescence of the H3 Ac-QTA**GG**STG-$NH_2$ control peptide.

## F. Use of small molecule post-translational modification receptors in peptide microarrays

### i. Background and Significance

The small molecules described above are quite unusual as PTM binding domains, not only due to their small size in comparison to typical protein binders and antibodies, but largely due to their ability to bind in a non-sequence selective fashion. Consequently, these molecules have the potential to serve as novel tools in the identification of sites of lysine and arginine methylation in proteins that have yet to be identified. To achieve this goal, the ability of these receptors to identify known sites of methylation within proteins must first be demonstrated. Ideally a screening approach would be used in a high-

throughput manner, making use of peptide arrays to readily identify PTM containing histone tails that a specific small molecule receptor interacts with.

To investigate binding in a microarray format we have established a collaboration with Dr. Brian Strahl's group in the Department of Biochemistry and Biophysics in the UNC School of Medicine. With their knowledge in the field of histone peptide microarrays, we hope to develop and establish methodology for screening histone peptides for their affinity for small molecule receptors such as **$A_2B$** and **$A_2D$**. The Strahl lab is developing the technology for such microarrays, in which histone PTM peptides are spotted in a grid format on a glass plate (Figure 2.92). Various histone sequences are individually synthesized containing different known PTMs, including all methylation states of both lysine and arginine. Each peptide is appended with a terminal biotin, which can subsequently be conjugated to a glass plate via a biotin-streptavidin interaction. Upon incubation with a known binder for a specific PTM, such as a protein or antibody as is being investigated by the Strahl lab, or a small molecule receptor in our case, the relative affinity of that binder for all the PTM containing peptides can be evaluated. Typically, after incubation with an effector protein, for example, the plate is then incubated with a fluorophore-containing antibody which recognizes and binds any proteins bound to the plate via interaction with a PTM peptide. A primary and then fluorophore-containing secondary antibody is sometimes used, while alternatively the direct fluorescent labeling of the PTM receptor is also possible. After various washes the fluorescence intensity of the plate can be visualized and quantified to some degree at each spot. While a bright spot is indicative of a tightly bound receptor, no signal is expected for the PTMs which do not bind the given receptor. This allows for the general

evaluation of both PTM affinity and selectivity, and can be used to assess PTM binding in a sequence selective or non-selective fashion, as well as screening for the effect of additional nearby PTMs.

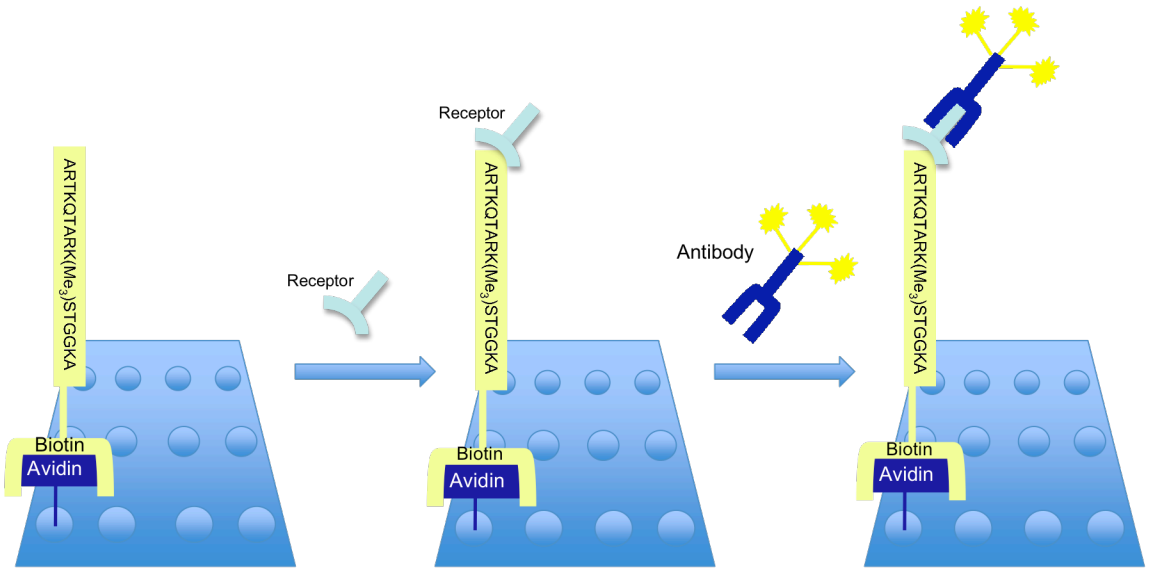
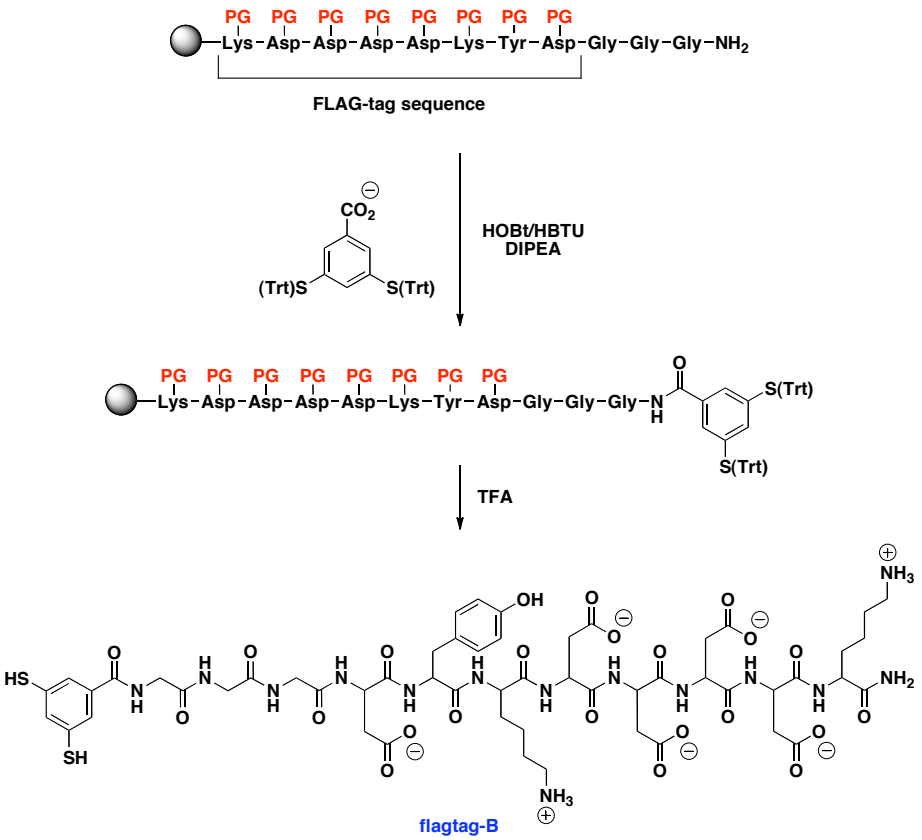



**Figure 2.92.** Peptide microarray assays. An array of histone peptides with desired modifications are conjugated onto the plate and then incubated first with the receptor of interest, followed by a fluorescently labeled antibody which recognizes the receptor and allows for visualization of the binding interaction.

While there is much to still be worked out in the implementation of this technology, it clearly has the potential to showcase the utility of small molecule PTM receptors. We hope to first demonstrate the affinity of $\mathbf{A_2B}$ for all KMe3 containing peptides, while also achieving a level of detection that clearly discriminates between KMe3 and the lower lysine methylation states. Eventually we hope to be able to use this microarray technology for the identification of unknown PTM sites by coating plates with, instead of synthetic well-characterized peptides, proteins with unknown PTMs.

### ii. Results, discussion, and ongoing work

**a. Generation 1 affinity labeled $A_2B$ receptor: $A_2$(B-flagtag).** To use **$A_2B$** as a PTM binder in the context of the peptide microarrays, the receptor must be selectively functionalized via synthetic modification with an affinity tag. We chose to append a FLAG-tag sequence to **$A_2B$**, as this peptide tag is utilized in many different assays that require recognition by an antibody. The peptide sequence of the FLAG-tag is N-DYKDDDDK-C. Due to the five carboxylates of **$A_2B$**, four of which are equivalent, the FLAG-tag peptide was first attached to dithiol **B** prior to equilibration with monomer **A**, so that the receptor would be selectively labeled at a single position. Dithiol monomer **flagtag-B** was synthesized via solid-phase synthesis of the eight residue peptide, with three glycine spacers to introduce some additional space between the unnatural receptor and the antibody recognition site (Scheme 2.6). The peptide was capped with a trityl-protected monomer **B** prior to deprotection and cleavage, and finally purified by HPLC.




**Scheme 2.6.** Synthesis of dithiol monomer **flagtag-B** via solid phase peptide synthesis.

Monomer **flagtag-B** was first investigated in the context of a DCC library. In the presence of two equivalents of **A** as well as a KMe3 dipeptide template, the FLAG-tag peptide had no discernable influence on the amplification of the **$A_2$(B-flagtag)** receptor (Figure 2.93). Both *rac*- and *meso*-**$A_2$(B-flagtag)** were amplified in a similar ratio as *rac*- and *meso*-**$A_2B$**. On a larger preparative scale, monomer **A** was equilibrated with **flagtag-B** in a biased library and templated with methylisoquinoline to generate higher yields of the **$A_2$(B-flagtag)** receptor than would otherwise be possible with KMe3, as shown in prior studies. Both *rac*-**$A_2$(B-flagtag)** and *meso*-**$A_2$(B-flagtag)** were well resolved and easily purified by semi-preparative HPLC. The ease of purification of the two isomers relative to **$A_2B$** is likely due to the increased hydrophilicity of the peptide-appended

receptor. Overall, this demonstrates the ease with which receptors identified by DCC can be synthetically modified, in contrast to traditional receptors.

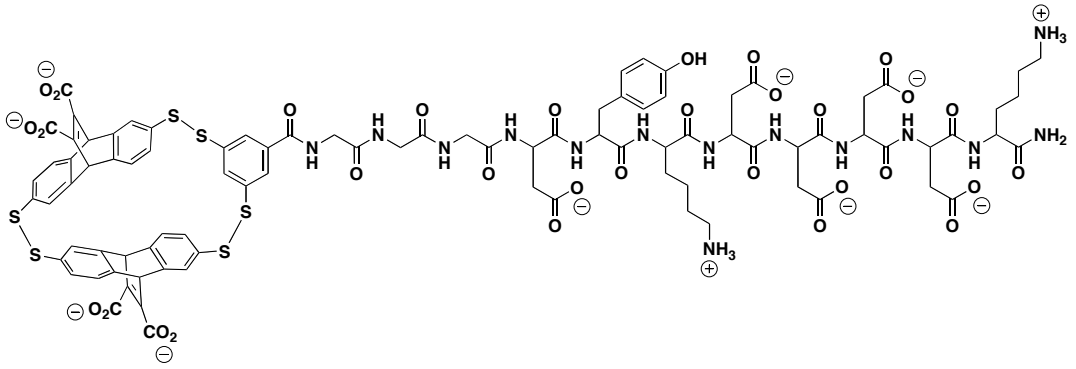


**Figure 2.93.** Structure of the affinity labeled **$A_2$(B-flagtag)** receptor.

**b. Microarray and binding studies with $A_2$(B-flagtag).** In an attempt to apply the designed **$A_2$(B-flagtag)** receptor to the peptide microarrays, the arrays were incubated with various concentrations of *rac*-**$A_2$(B-flagtag)** ranging from about 25 – 200 μM. After subsequent washes, incubation with the appropriate antibodies, and visualization, there was no evidence that **$A_2$(B-flagtag)** was bound to any PTMs on the array in these initial attempts. We postulated various reasons as to why this may have been the case. First, only three glycine residues separate the FLAG-tag antibody recognition site from the very unnatural **$A_2$B** aromatic receptor. It is possible that these two domains are in too close proximity, and therefore the presence of **$A_2$B** disturbs the interaction of the FLAG-tag antibody with its target sequence, preventing any visualization of **$A_2$(B-flagtag)** binding. Second, the role of the carboxylic acid of monomer **B** in the binding of **$A_2$B** to KMe3 and other methylated PTMs has not been investigated, and hence it is not known if this free carboxylate contributes to binding. It is suspected that monomer **B** is able to rotate within the context of the receptor, potentially allowing the carboxylate to interact

with the cationic PTMs. By appending a peptide at this site, this potential beneficial electrostatic interaction is no longer feasible, possibly weakening or preventing binding to the PTMs. Third, the *rac*-**$A_2$(B-flagtag)** solution was prepared in 10 mM pH 8.5 phosphate buffer, and the same buffer was used for subsequence plate washes throughout the experiment. While pH 8.5 buffer was used for consistency with prior binding studies, there is some question as to the stability of the FLAG-tag antibodies in basic solution. If the antibodies are consequently destabilized or unable to bind the **$A_2$(B-flagtag)** receptors interacting with the plate, this will again prevent any visualization of the PTM-receptor binding interaction. Last, it is also possible that the binding affinity of **$A_2$B** for KMe3 is relatively weak for these assays, and due to the number of successive steps and washes the **$A_2$(B-flagtag)** receptor is being washed off the plate.

To investigate the effect of attaching the FLAG-tag peptide to monomer **B** on binding, the affinity of *rac*-**$A_2$(B-flagtag)** for an H3 KMe3 peptide was investigated by fluorescence anisotropy. A similar eight residue H3 KMe3 peptide appended with an N-terminal fluorescein was investigated as in prior **$A_2$B** binding studies. The KMe3 peptide was found to bind *rac*-**$A_2$(B-flagtag)** with a dissociation constant of 52 μM, indicating that the FLAG-tag peptide appears to only have a minor influence of binding. This small difference in affinity between *rac*-**$A_2$(B-flagtag)** and *rac*-**$A_2$B** for KMe3 should not alone radically alter the results of the peptide microarrays. However, it must noted that the binding affinity of *rac*-**$A_2$(B-flagtag)** for KMe3 was measured under slightly different conditions and on a different instrument than prior *rac*-**$A_2$B** binding studies. Therefore these two dissociation constants should not necessarily be compared directly, and any discrepancy may purely be the result of experimental differences.

To address the issue of antibody recognition of *rac*-**$A_2$(B-flagtag)**, the fluorescence anisotropy experiment was repeated in the presence of the FLAG-tag antibody. Due to the extremely large size of the antibody in comparison to *rac*-**$A_2$(B-flagtag)**, it was expected that upon binding of the antibody bound *rac*-**$A_2$(B-flagtag)** to the KMe3 peptide, the change in fluorescence anisotropy would be much greater than without the antibody. This is due to the fact that the tumbling of the fluorophore should be more drastically altered when the peptide is bound to a large antibody-containing complex. However, this increase in fluorescence anisotropy signal was not observed upon introduction of the antibody, and closely mimicked the extent of change observed in the prior experiment. This indicated that the recognition of the FLAG-tag sequence by its appropriate antibody was likely not occurring, potentially due to the close proximity of the **$A_2$B** receptor. As expected, the presence of the antibody did not affect the measured dissociation constant ($K_d$ = 55 μM).

**c. Generation 2 affinity labeled $A_2$B receptor: (A-flagtag$_2$)$_2$B.** Based on the prior observations that the FLAG-tag antibody did not appear to be recognizing **$A_2$(B-flagtag)** receptor, that slightly weaker affinity for KMe3 resulted from modification of the carboxylate on monomer **B**, and that no binding was visualized in the initial peptide microarray experiments, the affinity labeled **$A_2$B** receptor was redesigned. First, the length of the spacer between the **$A_2$B** macrocycle and the FLAG-tag sequence was increased. Four glycine residues were included after the FLAG-tag sequence as opposed to three, followed by a diethylene glycol ($(PEG)_2$) spacer, lengthening the total distance between the two moieties by an additional 16 atoms. In addition, the FLAG-tag sequence was appended to monomer **A** instead of monomer **B**. The carboxylates on monomer **A**

are in fixed positions on the outside of the receptor, and therefore the mutation of these functional groups should have no influence on the receptor's affinity for methylated PTMs. Due to the equivalency of all four monomer **A** carboxylates, this will also result in four FLAG-tag sequences, and potentially a greater fluorescent signal upon binding to fluorescent antibodies (Figure 2.94).

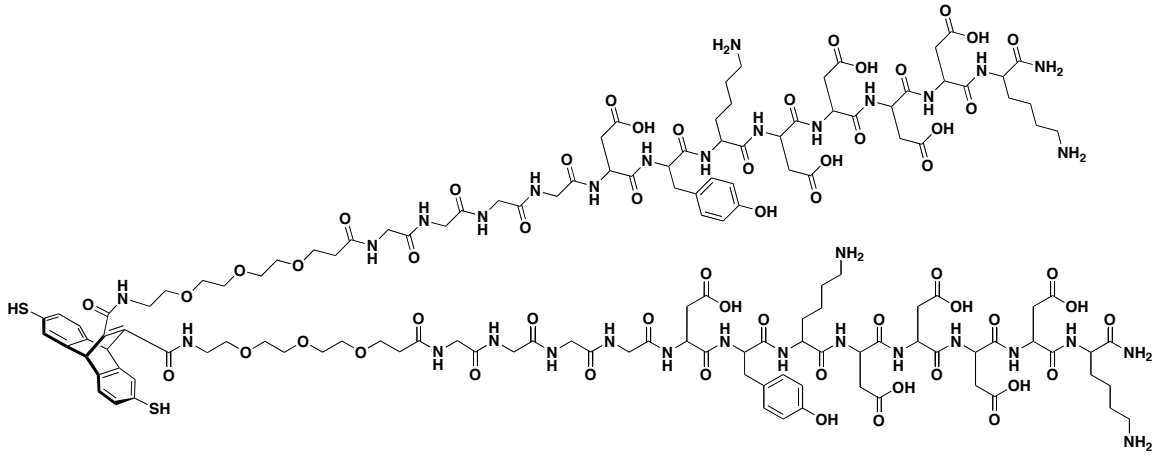


**Figure 2.94.** Structure of FLAG-tag labeled monomer **A** (**A-flagtag$_2$**).

The **(A-flagtag$_2$)$_2$B** receptor was prepared similarly to the first generation receptor. The peptides were capped with a trityl-protected monomer **A**, and it was found that the loading of the peptide synthesis resin was high enough to allow for each **A** to couple to two peptides. Upon cleavage and purification, monomer **A-flagtag$_2$** was equilibrated with monomer **B** in a biased library and templated with methylisoquinoline. Frequent LC/MS analysis revealed a slower than usual equilibration for this library, indicating that the presence of the appended peptides may be affecting the mobility of the monomers. After a month, the **(A-flagtag$_2$)$_2$B** receptors were purified on an analytical HPLC as a result of the low volume of the biased library (Figure 2.95). Due to extremely limited

amounts of material, the binding affinity of this receptor for KMe3 has not yet been investigated by fluorescence anisotropy.

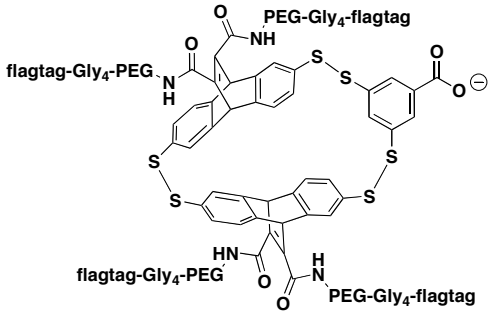



**Figure 2.95.** Structure of affinity labeled **(A-flagtag$_2$)$_2$B** receptor.

Future work involves testing this newly designed affinity tagged **$A_2$B** receptor in the peptide microarrays. A close to neutral pH buffer will be used for these experiments to ensure stability of the antibodies, and the FLAG-tag peptide should allow the receptor to solubilize at neutral pH. In addition, attempts to directly label the receptor with a fluorophore are underway to significantly cut back the number of steps prior to detection, diminishing the possibility that the receptor will be washed off following binding. In this case, the FLAG-tag sequence is no longer required for antibody recognition, and therefore fluorophores can be attached via the two lysine residues of the FLAG-tag sequence through an amide bond. This fluorophore-labeled receptor will also be investigated for its ability to selectively identify KMe3 containing peptides in the microarray assay.

### iii. Experimental Section.

**a. Synthesis of $A_2$(B-flagtag).** Synthesis of the FLAG-tag peptide was performed manually using Peptides International CLEAR-Amide resin. The peptide was

synthesized on a 0.2 mmol scale by standard Fmoc solid phase synthesis, with HOBt/HBTU coupling reagents in DMF. Following the synthesis of the FLAG-tag sequence, three glycine residues were coupled by standard methods. The peptide was capped with 1.4 equivalents of trityl-protected **B** (0.134 g, 0.28 mmol) and coupled overnight. Trityl protected monomer **B** was prepared previously by standard thiol protection methods.[60] Monomer **B** (1.8 g, 9.7 mmol) was suspended in 60 mL of TFA with Trt-OH (2.5 g, 9.7 mmol) and stirred under nitrogen for 1 hour. The TFA was evaporated and after addition of EtOAc, the product was extracted into 3M NaOH and then acidified with 10% HCl. The precipitate was filtered to give a white solid (5.5 g, 85%) and used for solid phase peptide synthesis without further purification. Upon coupling of protected monomer **B**, the peptide was deprotected and cleaved from the resin with a cocktail of 95% TFA/1.0% triisopropylsilane/2.0% $H_2O$/2.0% ethanedithiol (EDT) for 3 hours. The peptide was purified by semipreparative reverse-phase HPLC on a C18 column set to 35 °C and at a flow rate of 4 mL/min with a linear gradient of 0-60% B in 40 minutes. Once purified, the peptide was lyophilized to powder and characterized by ESI-MS ($[M+H]^+$ = 1352.3).

An **$A_2$(B-flagtag)** biased library was prepared on a 3.2 mL scale (**A**: 6.67 mM, 7.5 mg, 0.21 mmol; **flagtag-B**: 3.33 mM, 14.3 mg, 0.11 mmol) templated with methylisoquinoline triflate (10 mM, 9.3 mg, 0.32 mmol). The three solids were combined in a small vial and the pH of the solution was adjusted to 8.5 with 1M NaOH. Although some fine precipitate was initially visible, upon stirring the solution overnight the library solution turned completely soluble. After 5 days of stirring, the receptors were isolated by semi-preparative HPLC. Approximately 0.4 mL injections were

chromatographed using standard peptide synthesis mobile phases A and B (0-80% B 0-40 min) with a flow rate of 4.0 mL/min and a column temperature of 40 °C. The two **$A_2$(B-flagtag)** peaks at 22.5 and 23 minutes were collected separately (Figure 2.96) and analyzed for purity by analytical LC-MS. Both peaks are indistinguishable by mass ($[M+H]^+$ = 2057.3, $[M+2H]^{2+}$ = 1029.1). Purified *rac*- and *meso*-**$A_2$(B-flagtag)** were lyophilized to powder and stored under nitrogen.

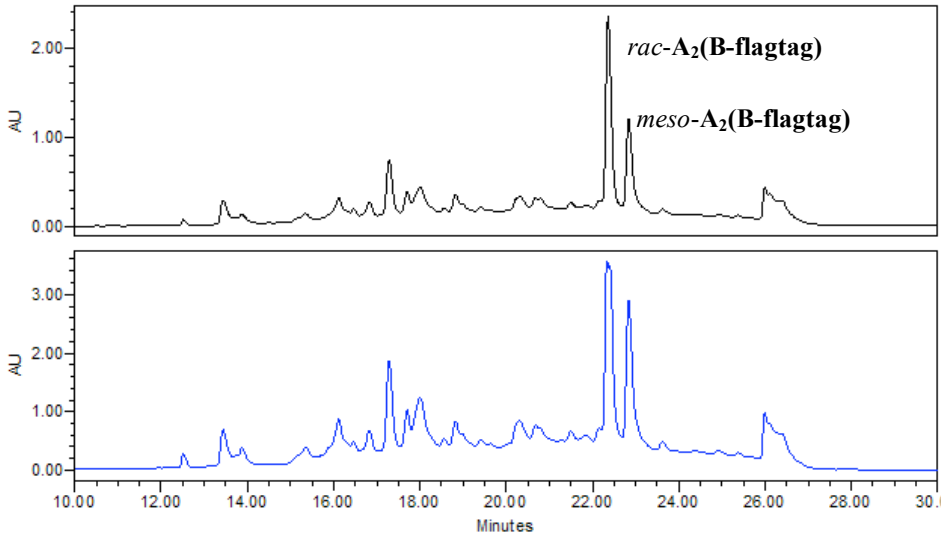



**Figure 2.96.** Semi-preparative HPLC trace of an **$A_2$(B-flagtag)** biased library (black = 254 nm, blue = 214 nm).

**b. Fluorescence anisotropy binding experiments.** Binding assays were performed with purified *rac*-**$A_2$(B-flagtag)** and a fluorescein labeled histone 3 K9Me3 peptide. The peptide was dissolved in buffer (10 mM phosphate pH 8.5), and a 25 μM solution was prepared as determined by UV-Vis at 492 nm (ε = 78,000 $M^{-1}cm^{-1}$). A stock solution of *rac*-**$A_2$(B-flagtag)** was prepared in identical buffer, with the solution concentration determined by mass. Individual solutions were prepared for each concentration of *rac*-**$A_2$(B-flagtag)** in the first experiment, whereas serial dilutions were used to prepare each concentration of *rac*-**$A_2$(B-flagtag)** in the second experiment.

Preparation by serial dilutions of the *rac*-**$A_2$(B-flagtag)** + antibody samples allowed for the ratio of *rac*-**$A_2$(B-flagtag)** : antibody to remain constant throughout the experiment. Assays were prepared in 96-well half-area plates (Corning) with a total volume of 50 μL per well, containing 5 μM labeled peptide and increasing concentrations of *rac*-**$A_2$(B-flagtag)** (with or without the FLAG-tag antibody) in buffer. Plates were spun down and allowed to incubate for at least 30 minutes before analysis. Fluorescence anisotropy was measured on a POLARstar plate reader (BMG Labtech) using FP485, 520A, and 520B filters at approximately 25°C. The anisotropy data was plotted as a function **$A_2$B** concentration and each plot was fitted in KaleidaGraph to the following equation:

$$r = \left( \left( \frac{(a + x + k_d) \pm \sqrt{(-a - x - k_d)^2 - 4(a \cdot x)}}{2 \cdot a} \right) \cdot (r_\infty - r_o) \right) + r_o$$

where r is fluorescence anisotropy, $r_o$ is the initial anisotropy value, $r_\infty$ is the maximum anisotropy value, a is the peptide concentration, x is the concentration of **$A_2$B**, and $k_d$ is the dissociation constant. These measurements were taken in only once.

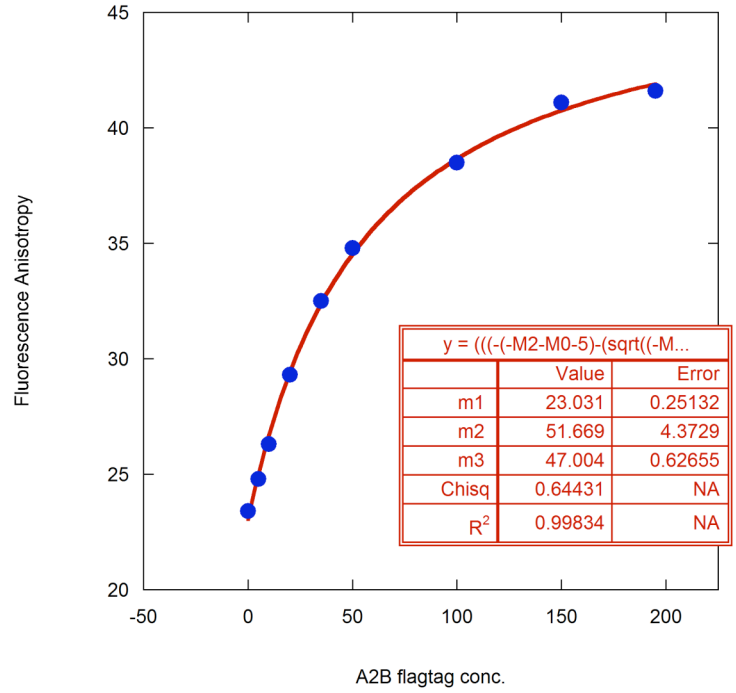



**Figure 2.97.** Fluorescence anisotropy of *rac*-**$A_2$(B-flagtag)** with H3 $KMe_3$ (m2 = dissociation constant).

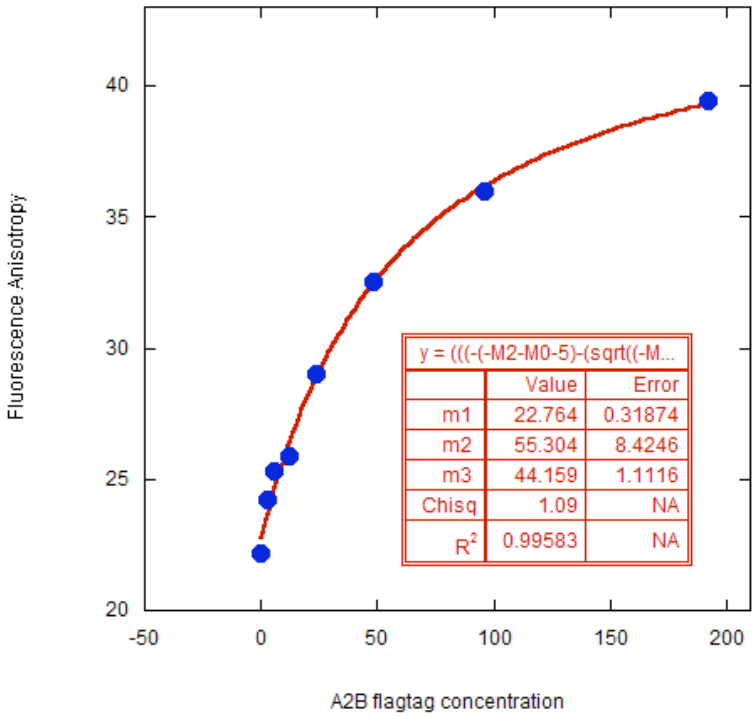



**Figure 2.98.** Fluorescence anisotropy of *rac*-**$A_2$(B-flagtag)** with H3 $KMe_3$, in the presence of the FLAG-tag antibody (m2 = dissociation constant).

**c. Synthesis of (A-flagtag$_2$)$_2$B.** Synthesis of the FLAG-tag peptide was performed by hand on a 0.1 mmol scale as discussed previously, followed by the coupling of four glycine residues. Two equivalents of Fmoc-NH-(PEG)$_2$-CO$_2$H (0.089 g, 0.2 mmol) purchased from Novabiochem were coupled overnight to introduce the PEG spacer. The peptide was capped with about 0.2 mmol of trityl-protected **A** (0.168 g) and coupled overnight. Trityl protected monomer **A** was prepared previously by standard thiol protection methods with Trt-Cl.[61] Monomer **A** (0.11 g, 0.3 mmol) was added to trityl chloride (0.18g, 0.65 mmol) in dichloromethane (1.5 mL) and stirred for three hours. The reaction was quenched with 1 mL of water and a precipitate was formed. The water layer was discarded followed by evaporation of the dichloromethane, and any residual water was removed by lyophilization. The resulting tan solid (0.24 g, quant.) was used for solid phase peptide synthesis without further purification. Upon capping with protected **A**, the peptide was deprotected and cleaved from the resin with a cocktail of 95% TFA/1.0% triisopropylsilane/2.0% $H_2O$/2.0% ethanedithiol (EDT) for 3 hours. The peptide was purified by semipreparative reverse-phase HPLC on a C18 column set to 35 °C and at a flow rate of 4 mL/min with a linear gradient of 0-60% B in 40 minutes. Efforts must be taken to avoid disulfide exchange during purification by keeping the crude solution in aqueous acidic media. In this case, the reducing agent *tris*(2-carboxyethyl)phosphine (TCEP) was added to the crude peptide solution to reduce any macrocycles formed. Once purified, the peptide was lyophilized to powder and characterized by ESI-MS ($[M+2H]^{2+}$ = 1603.6, $[M+3H]^{3+}$ = 1069.4, $[M+4H]^{4+}$ = 802.3, $[M+5H]^{5+}$ = 642.0). This synthesis afforded an extremely low yield.

A **(A-flagtag$_2$)$_2$B** biased library was prepared on a 95 µl scale (**A-flagtag$_2$**: 5 mM, 1.5 mg, 0.48 µmol; **B**: 2.5 mM, 0.044 mg, 0.24 µmol) templated with methylisoquinoline triflate (7.5 mM, 0.21 mg, 0.72 µmol). Monomer **A-flagtag$_2$** was dissolved directly into 50 mM sodium phosphate buffer pH 8.5, and aliquots of **B** and methylisoquinoline were added from larger stock solutions to achieve the desired concentrations. The solution sat in a capped vial and was frequently analyzed by LC/MS (0.2% formic acid aqueous and acetonitrile mobile phases; 10-22% B 0-3 minutes, then 22-55% B 3-17 minutes). After over a month of equilibration, the receptors were isolated via automated 6-8 µl injections on a Waters Alliance analytical HPLC. The library was chromatographed using an Atlantis C18 column (4.6 × 150 mm, 5 µm) with a flow rate of 1.0 mL/min and a column temperature of 35 °C (0.1% TFA aqueous and acetonitrile mobile phases; 10-39% B 1-22 minutes). Poor resolution of the two isomers was achieved, however the two **(A-flagtag$_2$)$_2$B** peaks at 18 minutes were collected separately (figure 2.98) and analyzed for purity by analytical LC-MS. Both peaks are indistinguishable by mass ($[M+5H]^{5+}$ = 1319.7, $[M+6H]^{6+}$ = 1099.9, $[M+7H]^{7+}$ = 942.9, $[M+8]^{8+}$ = 825.2). Purified **(A-flagtag$_2$)$_2$B** was lyophilized to powder.

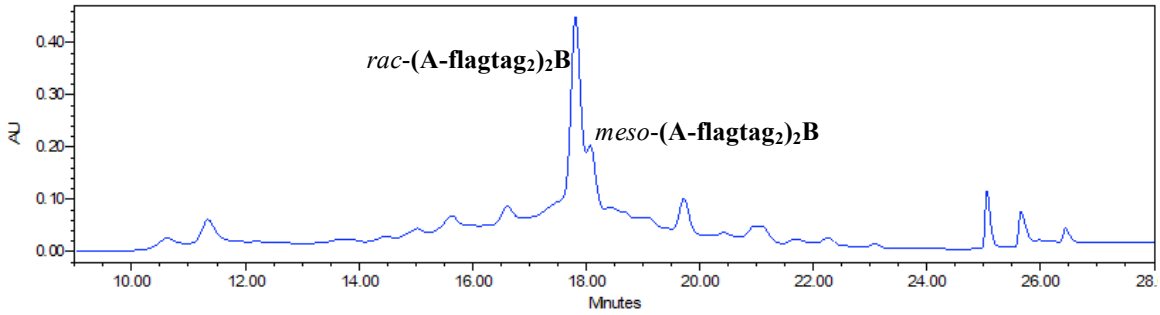



**Figure 2.99.** Semi-preparative HPLC trace of an **(A-flagtag$_2$)$_2$B** biased library (black = 214 nm, blue = 254 nm).

[1] (a) Bhaumik, S. R.; Smith, E.; Shilatifard, A. *Nature Struct. Molec. Biol.* **2007**, *14*,1008-1016. (b) Wang, G.G.; Allis, C.D.; Chi, P. *Trends in Molec. Med.* **2007**, *13*, 363-372.

[2] Zhang, Y.; Reinberg, D. *Genes & Development* **2001**, *15*, 2343-2360.

[3] Arents, G.; Burlingame, R. W.; Wang, B. C.; Love, W. E.; Moudrianakis, E. N. *Proc.* Natl. Acad. Sci. **1991**, 88, 10148-10152.

[4] Khorasanizedeh, S. *Cell* **2004**, *116*, 259-272.

[5] (a) Jenuwein, T.; Allis, C. D. *Science* **2001**, *293*, 1074-1080. (b) Kouzarides, T. *Cell* **2007**, *128*, 693-705.

[6] Berger, S. L. *Nature* **2007**, *447*, 407-412.

[7] Martin, C.; Zhang, Y. *Nature Rev. Molec. Cell Biol.* **2005**, *6*,838-849.

[8] Strahl, B. D.; Allis, C. D. *Nature* **2000**, *403*, 41-45.

[9] (a) Fischle, W.; Tseng, B. S.; Dormann, H. L.; Ueberheide, B. M.; Garcia, B. A.; Shabanowitz, J.; Hunt, D. F.; Funabiki, H.; Allis, C. D. *Nature* **2005**, *438*, 1116-1122. (b) Hirota, T.; Lipp, J. J.; Toh, B. H.; Peters, J. M. *Nature* **2005**, *438*, 1176-1180.

[10] (a) Seet, B. T.; Dikic, I.; Zhou, M. M.; Pawson, T. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 473-483. (b) Ruthernburg, A. J.; Allis, C. D.; Wysocka, J. *Cell* **2007**, *25*, 15-30.

[11] (a) Bottomley, M. J. *Embo Reports* **2004**, *5*, 464-469. (b) Daniel, J. A.; Pray-Grant, M. G.; Grant, P. A. *Cell Cycle* **2005**, *4*, 919-926.

[12] Taverna, S. D.; Li, Haitao, Ruthenburg, A. J.; Allis, C. D.; Patel, D. J. *Nat. Struct. Mol. Biol.* **2007**, *14*, 1025-1040.

[13] (a) Ma, J. C.; Dougherty, D. A. *Chem. Rev.* **1997**, *97*, 1303-1324. (b) Gallivan, J. P.; Dougherty, D. A. *Proc. Nat. Acad. Sci.* **1999**, *96*, 9459-9464.

[14] Hughes, R. M.; Wiggins, K. R.; Khorasanizadeh, S.; Waters, M. L. *Proc. Nat. Acad. Sci. USA* **2007**, *104*, 11184-11188.

[15] Li, H.; Fischle, W.; Wang, W.; Duncan, E. M.; Liang, L.; Murakami-Ishibe, S.; Allis, C. D.; Patel, D. J. *Mol. Cell* **2007**, *28*, 667-691.

[16] Li, H. T.; Ilin, S.; Wang, W. K.; Duncan, E. M.; Wysocka, J.; Allis, C. D.; Patel, D. J. *Nature* **2006**, *442*, 91-95.

[17] Jacobs, S. A.; Khorasanizadeh, S. *Science* **2002**, *295*, 2080-2083.

[18] (a) Nielsen, A. L.; Oulad-Abdelghani-M.; Ortiz, J. A.; Remboutsika, E.; Chambon, P.; Losson, R. *Mol. Cell* **2001**, *7*, 729-739. (b) Zhao, T.; Heyduk, T.; Allis, C. D.; Eissenberg, J. C. *J. Biol. Chem.* **2000**, *275*, 28332-28338.

[19] Nielsen, P. R.; Nietlispach, D.; Mott, H. R.; Callaghan, J.; Bannister, A.; Kouzarides, T.; Murzin, A. G.; Murzina, N. V.; Laue, E. D. *Nature* **2002**, *416*, 103-107.

[20] Bannister, A. J.; Kouzarides, T. *Nature* **2005**, *436*, 1103-1106.

[21] (a) McBride, A. E.; Silver, P. A. *Cell* **2001**, *106*, 5-8. (b) Lee, D. Y.; Teyessier, C.; Strahl, B. D.; Stallcup, M. R. *Endocrine Reviews* **2005**, *26*, 147-170. (c) Bedford, M. T.; Richard, S. *Molecular Cell* **2005**, *18*, 263-272.

[22] Cosgrove, M. S.; Boeke, J. D.; Wolberger, C. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1037-1043.

[23] (a) Burley, S. K.; Petska, G. A. *FEBS Letters* **1986**, *203*, 139-143. (b) Flocco, M. M.; Mowbray, S. L. *J. Mol. Biol.* **1994**, *235*, 709-717. (c) Mitchell, J. B. O.; Nandi, C. L.; McDonald, I. K.; Thornton, J. M.; Price, S. L. *J. Mol. Biol.* **1994**, *239*, 315-331.

[24] Hughes, R. M.; Waters, M. L. *J. Am. Chem. Soc.* **2006**, *128*, 12735-12742.

[25] (a) Sprangers, R.; Groves, M. R.; Sinning, I.; Sattler, M. *J. Mol. Biol.* **2003**, *327*, 507-520. (b) Cote, J.; Richard, S. *J. Biol. Chem.* **2005**, *280*, 28476-28483.

[26] Wang, H.; Huang, Z. Q.; Xia, L.; Feng, Q.; Erdjument-Bromage, H.; Strahl, B. D.; Briggs, S. D.; Allis, C. D.; Wong, J.; Tempst, P.; Zhang, Y. *Science* **2001**, *293*, 853-857.

[27] (a) Kirmizis, A.; Santos-Rosa, H.; Penkett, C. J.; Singer, M. A.; Verrmeulen, M.; Mann, M.; Bahler, J. Green, R. D.; Kouzarides, T. *Nature* **2007**, *449*, 928-932. (b) Guccione, E.; Bassi, C.; Casadio, F.; Martinato, F.; Cesaroni, M.; Schuchlautz, H.; Luscher, B.; Amati, B. *Nature* **2007**, *449*, 933-937.

[28] (a) Blow, N. *Nature* **2007**, *447*, 741–744. (b) Kazanecki, C. C.; Kowalski, A. J.; Ding, T.; Rittling, S. R.; Denhardt, D. T. *J. Cell. Biochem.* **2007**, *102*, 925–935.

[29] Taverna, S. D.; Ueberheide, B. M.; Liu, Y.; Tackett, A. J.; Diaz, R. L.; Shabanowitz, J.; Chait, B. T.; Hunt, D. F.; Allis, C. D. *Proc. Nat. Acad. Sci.* **2007**, *104*, 2086-2091.

[30] Jensen, O. N. *Nature Rev. Mol. Cell Biol.* **2006**, *7*, 391-403.

[31] Freitas, M. A.; Sklenar, A. R.; Parthun, M. R. *J. Cell Biochem.* **2004**, *92*, 691-700.

[32] Zhang, K.; Siino, J. S.; Jones, P. R.; Yau, P. M.; Bradbury, E. M. *Proteomics* **2004**, *4*, 3765-3774.

[33] Zubarev, R. A. *Curr. Opin. Biotechnol.* **2004**, *15*, 12-16.

[34] (a) Pesavento, J. J.; Kim, Y. B.; Taylor, G. K.; Kelleher, N. L. *J. Am. Chem. Soc.* **2004**, *126*, 3386-3387. (b) Boyne, M. T.; Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L. *J. Proteome Res.* **2006**, *5*, 248-253. (c) Thomas, C. E; Kelleher, N. L.; Mizzen, C. A. *J. Proteome Res.* **2006**, *5*, 240-247. (d) Siuti, N.; Roth, M. J.; Mizzen, C. A.; Kelleher, N. L. Pesavento, J. J. *J. Proteome Res.* **2006**, *5*, 233-239.

[35] Williams, B. A. R.; Lin, L. Lindsay, S. M.; Chaput, J. C. *J. Am. Chem. Soc.* **2009**, *106*, 6330-6331.

[36] Beshara, C. S.; Jones, C. E.; Daze, K. D.; Lilgert, B. J.; Hof, F. *ChemBioChem* **2010**, *11*, 63-66.

[37] For reviews, see: (a) Corbett, P. T.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J-L.; Sanders, J. K. M.; Otto, S. *Chem. Rev.* **2006**, *106*, 3652-3711. (b) Lehn, J-M. *Chem. Eur. J.* **1999**, *5*, 2455-2463. (c) Ladame, S. *Org. Biomol. Chem.* **2008**, *6*, 219-226. (d) Ludlow, R. F.; Otto, S. *Chem. Soc. Rev.* **2008**, *37*, 101-108. (e) Otto, S. *Curr. Opin. Drug Disc. Devel.* **2003**, *6*, 509-520.

[38] (a) Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *J. Am. Chem. Soc.* **2000**, *122*, 12063-12064. (b) Ramström, O.; Lehn, J.-M. *ChemBioChem* **2000**, *1*, 41-48.

[39] (a) Kearney, P. C.; Mizoue, L. S.; Kumpf, R. A.; Forman, J. E.; McCurdy, A.; Dougherty, D. A. *J. Am. Chem. Soc.* **1993**, *115*, 9907-9919. (b) Stauffer, D. A.; Barrans, R. E.; Dougherty, D. A. *J. Org. Chem.* **1990**, *55*, 2762-2767.

[40] (a) Ngola, S. M.; Kearney, P. C.; Mecozzi, S.; Russell, K.; Dougherty, D. A. *J. Am. Chem. Soc.* **1999**, *121*, 1192-1201. (b) Petti, M. A.; Shepodd, T. J.; Barrans, R. E.; Dougherty, D. A. *J. Am. Chem. Soc.* **1988**, *110*, 6825-6840.

[41] Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *Science* **2002**, *297*, 590-593.

[42] Corbett, P. T.; Tong, L. H.; Sanders, J. K. M.; Otto, S. *J. Am. Chem. Soc.* **2005**, *127*, 8902-8903.

[43] Corbett, P. T.; Sanders, J. K. M.; Otto, S. *Chem. Eur. J.* **2008**, *14*, 2153-2166.

[44] (a) Field, L.; Engelhardt, P. R. *J. Org. Chem.* **1970**, *35*, 3647-3655. (b) Staab, H. A.; Kirrstetter, R. G. H. *Liebigs Ann. Chem.* **1979**, 886.

[45] Kretsinger, J. K.; Schneider, J. P. *J. Am. Chem. Soc.* **2003**, *125*, 7907-7913.

[46] (a) Hughes, R. M.; Benshoff, M. L.; Waters, M. L. *Chem. Eur. J.* **2007**, *13*, 5753-5764. (b) Salonen, L. M.; Bucher, C.; Banner, D. W.; Haap, W.; Mary, J.-L.; Benz, J.; Kuster, O.; Seiler, P. I.; Schweizer, W. B.; Diederich, F. *Angew. Chem. Int. Ed.* **2009**, *48*, 811-814. (c) Whiting, A. L.; Neufeld, N. M.; Hof, F. *Tetrahedron Lett.* **2009**, *50*, 7035-7037.

[47] Sjoback, R.; Nygren, J.; Kubista, M. *Spectrochimica Acta A* **1995**, *51*, L7-L21.

[48] Gelman, M. A.; Richter, S.; Cao, H.; Umezawa, N.; Gellman, S. H.; Rana, T. M. *Org. Lett.* **2003**, *5*, 3563-3565.

[49] Suksai, C.; Gomez, S. F.; Chhabra, A.; Liu, J.; Skepper, J. N.; Tuntulani, T.; Otto, S. *Langmuir* **2006**, *22*, 5994-5997.

[50] Huang, J.; Berger, S. L. *Curr. Opin. Genet. Dev.* **2008**, *18*, 152-158.

[51] Liénard, B. M. R.; Selevsek, N.; Oldham, N. J.; Schofield, C. J. *ChemMedChem* **2007**, *2*, 175-179.

[52] Wüthrich, K. *NMR of Proteins and Nucleic Acids*. John Wiley and Sons, Inc.: 1986.

[53] Wang, Y.; Killian, J.; Hamasaki, K.; Rando, R. R. *Biochemistry* **1996**, *35*, 12338-12346.

[54] West, K. R.; Ludlow, R. F.; Corbett, P. T.; Besenius, P.; Mansfeld, F. M.; Cormack, P. A. G.; Sherrington, D. C.; Goodman, J. M.; Stuart, M. C. A.; Otto, S. *J. Am. Chem. Soc.* **2008**, *130*, 10834-10835.

[55] Hughes, R. M.; Waters, M. L. *J. Am. Chem. Soc.* **2005**, *127*, 6518-6519.

[56] Yana, D.; Shimizu, T.; Hamasaki, K.; Mihara, H.; Ueno, A. *Macromol. Rapid Commun.* **2002**, *23*, 11-15.

[57] Corbett, P. T.; Sanders, J. K. M.; Otto, S. *J. Am. Chem. Soc.* **2005**, *127*, 9390-9392.

[58] (a) Grote, Z.; Scopelliti, R.; Severin, K.. *Angew. Chem. Int. Ed.* **2003**, *42*, 3821-3825. (b) Severin, K. *Chem. Eur. J.* **2004**, *10*, 2565-2580.

[59] Corbett, P. T.; Sanders, J. K. M.; Otto, S. *Angew. Chem. Int. Ed.* **2007**, *46*, 8858-8861.

[60] Fernandes, C.; Correia, J. D. G.; Gano, L.; Santos, I.; Seifert, S.; Syhre, R.; Bergmann, R.; Spies, H. *Bioconjugate Chem.* **2005**, *16*, 660-668.

[61] Bang, D.; Pentelute, B. L.; Gates, Z. P.; Kent, S. B. *Org. Lett.* **2006**, *8*, 1049-1052.

# CHAPTER III

# PHOTOSWITCHABLE DYNAMIC COMBINATORIAL LIBRARIES

(Reproduced, in part, with permission from Ingerman, L. A.; Waters, M. L. *J. Org. Chem*. **2009**, *74*, 111 – 117.)

## A. Background and Significance

With the establishment of DCC as an effective method for the identification of novel receptors based on simple building blocks, there is much interest in expanding the scope of receptors than can be generated in such dynamic libraries. In particular, incorporating building blocks with new features and potentially interesting recognition elements is desirable in order to expand the present applications of DCC. We have focused our efforts in this area on the development of multilevel dynamic libraries by combining two reversible processes, hydrazone exchange and photoinduced isomerization.

**i. Doubly dynamic libraries.** To further increase library diversity several groups have developed doubly dynamic DCLs through the incorporation of two orthogonal reversible reactions which can be triggered independently, demonstrating the benefit of incorporating multiple equilibria in a single system.[1] Having two or more reversible chemistries that can be switched on and off independently allows for enhanced control of the system, while allowing for new potential applications by alternating use of different chemistries. Otto and co-workers have prepared DCLs which feature two simultaneous

covalent exchange reactions, disulfide and thioester exchange.[2] In this system the two reversible reactions are addressed sequentially. First, library equilibration occurs based on thioester exchange only, and then in the presence of atmospheric oxygen both reactions occur simultaneously. They later report the combination of hydrazone and disulfide exchange in a single system, which can be either operated fully orthogonally or occur simultaneously at any point depending on the pH of the solution.[3]

Eliseev and co-workers have explored doubly dynamic libraries as well by combining non-covalent metal coordination with imine exchange, which can be used as independent equilibrium processes controlled by different types of external intervention, oxidation/reduction of the metal template and change in the pH and temperature of the medium.[4] Others have since taken advantage of this methodology incorporating metal-ligand coordination and reversible covalent exchange for the development of hierarchically self-assembled systems including iron(II) and copper(I) complexes,[5] as well as zinc(II) grid complexes.[6] Furthermore, a system has been reported in which three dynamic linkages, disulfide, imine, and coordinative bonds, were shown to be capable of simultaneous reversible exchange.[7] Although the three types of dynamic linkages were demonstrated to be mutually compatible, both transmetallation and covalent imine exchange were used to alter the equilibrium between disulfides, allowing for greater control over the degree of self-sorting. The parallel utilization of three different reversible reactions has also been shown in the assembly of boronic acid based macrocycles and cages of unprecedented size in one-pot reactions.[8] In this case, aldehyde-amine condensation reactions and metal-ligand interactions are combined with the reversible formation of boronic esters from boronic acids and diols.

**ii. Azobenzene as an optical trigger.** Despite its extensive use in other applications, particularly in the field of molecular recognition, photochemistry has yet to be widely employed in the design of dynamic combinatorial libraries. However light offers many advantages as a means of manipulating systems with extreme spatial and temporal resolution, often reversibly. Eliseev and coworkers reported an early example which integrated photoisomerization into DCC, making use of an unsaturated dicarboxylate monomer in the development of anionic receptors for arginine.[9]

Azobenzene has been widely used as an optical trigger for various photoresponsive systems due to its pronounced changes in geometry and polarity upon light-induced isomerization. Azobenzene is an attractive photoswitch due to its high photo stability, facile isomerization resulting in good quantum yields, and extremely fast and reversible isomerization processes (picosecond time scale).[10] At thermal equilibrium the trans isomer is dominant, but irradiation to the photostationary state converts the trans isomer to its corresponding cis form (Figure 3.1). The reverse process is also feasible photochemically (at 450 nm) or thermally, although thermal relaxation to the trans state is a slow process (hour-to-day time scale).[11]

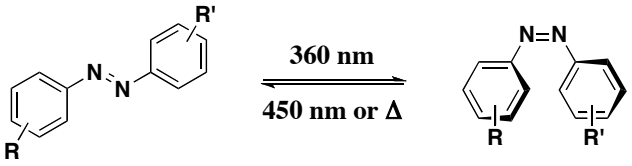



**Figure 3.1.** Cis-trans isomerization of azobenzene derivatives.

The conformational change that is induced upon isomerization of azobenzene derivatives has been successfully exploited to control the biological properties of various systems, such as folded peptides[12] and helical polymers,[13] by either disrupting, changing,

or enhancing secondary structure. Recently, progress has been made in the development of photoswitches that covalently modify target proteins and reversibly present and withdraw a ligand from its binding site as a result of photoisomerization of an azobenzene linker, allowing for rapid and selective manipulation of protein function.[14] The photoswitchable properties of azobenzenes have also been utilized to manipulate the properties of host-guest systems involving crown ethers[15] and cyclodextrins,[16] while also finding applications as small molecule inhibitors.[17]

**iii. Hydrazone exchange in dynamic combinatorial libraries.** Hydrazone exchange is well suited for DCC, as much success has been met with this reversible reaction in the preparation of DCLs.[18] The required hydrazide and aldehyde functionality can be incorporated in a straightforward manner, and the hydrazone linkage can be subsequently formed under acidic conditions (Figure 3.2). While acid catalyzes both the initial formation and the interconversion of an assembly of macrocycles, neutralization yields stable, isolable products.[19] The hydrazone exchange reaction is chemoselective and thus compatible with a wide range of solvents and functional groups. DCLs based on hydrazone exchange have been reported using metal ions, carbohydrates, nucleic acids, or peptides as either building blocks or templates without observing any interference.[20]

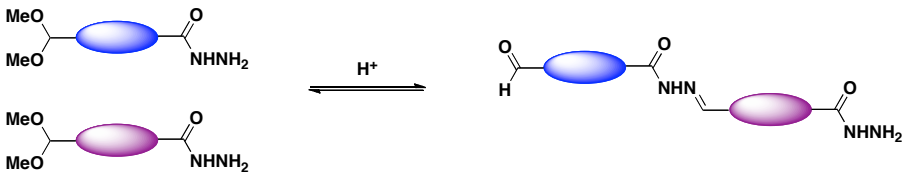
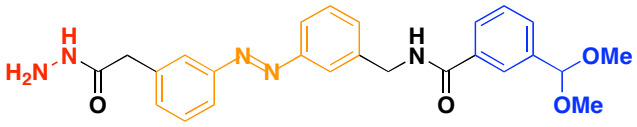



**Figure 3.2.** Hydrazone formation from hydrazine monomers.

**iv. Goal of this work.** In this work we aimed to both investigate the incorporation of a photochemical switch in a dynamic combinatorial library in attempt to develop novel molecular receptors, while also expanding the field of doubly dynamic libraries by combining hydrazone exchange and photoinduced isomerization in a single system. The two exchange processes involved in our double-level DCLs is advantageous in that it offers a higher degree of control over the library composition in the investigation of potential targets. While hydrazone exchange facilitates the traditional formation and interconversion of an assembly of macrocycles under acidic conditions, photoinduced isomerization can be applied for the development of switchable receptors. We aimed to develop a DCL from which we could identify switchable receptors with which one could photomodulate molecular recognition processes as a direct result of the distinct conformational changes of azobenzene.

## B. Results and Discussion

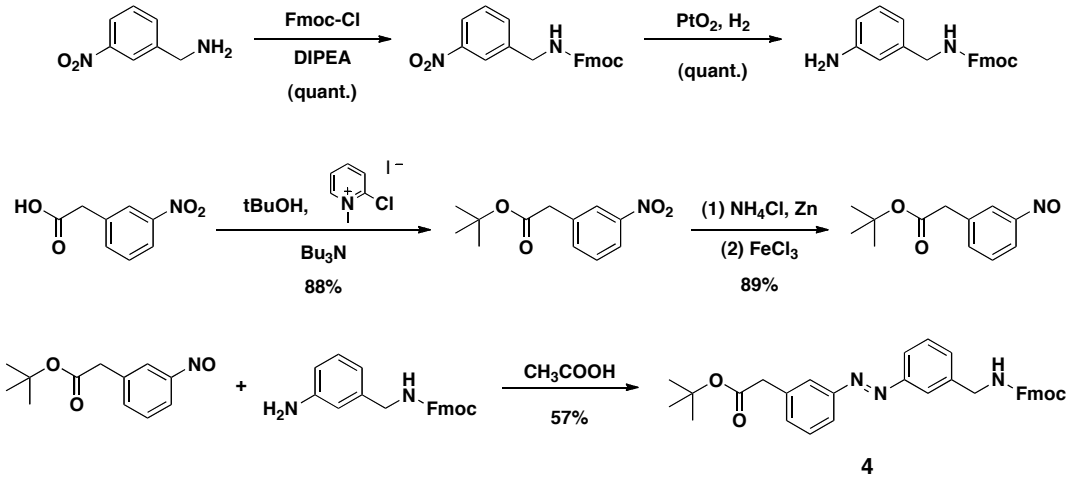
**i. Azobenzene monomer design and synthesis.** Building block **1** was designed containing the desired azobenzene chromophore, appended with the necessary hydrazide and protected aldehyde functionality to facilitate hydrazone exchange (Figure 3.3). The formation of an acyl hydrazone was desired, for without an acyl or similar electron withdrawing group hydrazones tend to be too stable for use in dynamic libraries. Furthermore, an aromatic acetal was chosen, as it has been observed that the equilibrium constant of hydrazone formation is several times higher with an aromatic aldehyde than with an aliphatic one.[21]



**1**

**Figure 3.3.** Azobenzene building block **1** functionalized with a hydrazide (red) and protected aldehyde (blue) to facilitate hydrazone exchange.

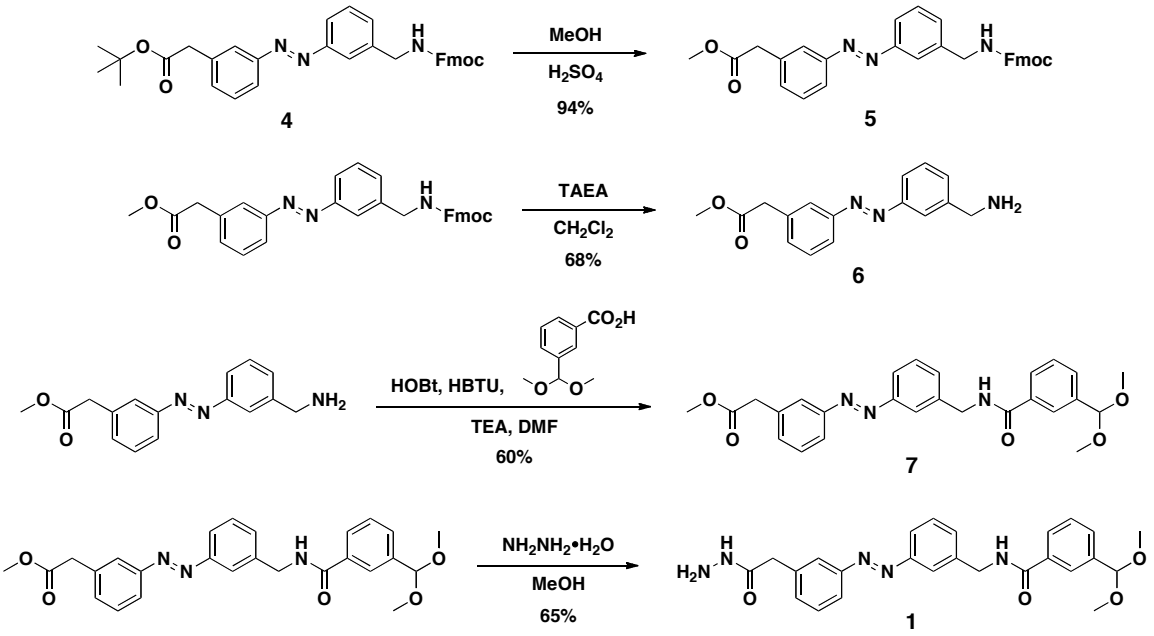
Monomer **1** was synthesized based on the corresponding fully protected azobenzene amino acid **4**. The preparation of **4** (Scheme 3.1) relies on the reaction of a nitrosobenzene with an aniline as previously described by Hilvert and coworkers.[22] The nitrosobenzene was prepared in two steps by esterification of commercially available *m*-nitrophenylacetic acid under Mukaiyama conditions, followed by reduction of the nitro group to the hydroxylamine and immediate reoxidation to the nitroso compound. The aniline was prepared in two steps from commercially available 3-nitrobenzylamine hydrochloride, involving introduction of an Fmoc protecting group followed by reduction of the nitro group. Coupling the nitrosobenzene and aniline in glacial acetic acid resulted in the reported protected amino acid **4**.



**4**

**Scheme 3.1**. Preparation of precursor **4**.[22]

*Tert*-butyl ester **4** was transesterified to the corresponding methyl ester **5** because we were unable to prepare **5** directly due to the instability of the required 3-nitrosophenylacetic acid methyl ester precursor (Scheme 3.2).[23] Fmoc deprotection under basic conditions yielded the free amine **6**. The coupling of **6** with 3-carboxybenzaldehyde dimethoxyacetal via standard HOBt/HBTU coupling to yield **7**, followed by hydrazinolysis of the methyl ester afforded the desired building block **1** as a mixture of cis and trans isomers.




**Scheme 3.2.** Synthesis of building block **1**.

**ii. Library formation under thermodynamic conditions.** Deprotection and subsequent cyclization of **1** (4 mM) was accomplished using an excess of TFA (100 mM). Because we were met with solubility problems upon cyclization in most solvents, libraries prepared with building block **1** only were generated in DMSO. In this case the entire library remained soluble throughout the time of analysis. Reactions were

monitored daily by LC-MS for 4-5 days, although thermodynamic equilibrium was reached after 4 days. Most cyclization occurred within 24 hours and the initial library distribution did not change drastically after the first day. Although linear species were observed within hours of monomer deprotection, the libraries were composed of entirely cyclic macrocycles upon reaching equilibrium. This DCL gave relatively simple distributions dominated by macrocyclic monomers and dimers, along with smaller trimer peaks (Figure 3.4).

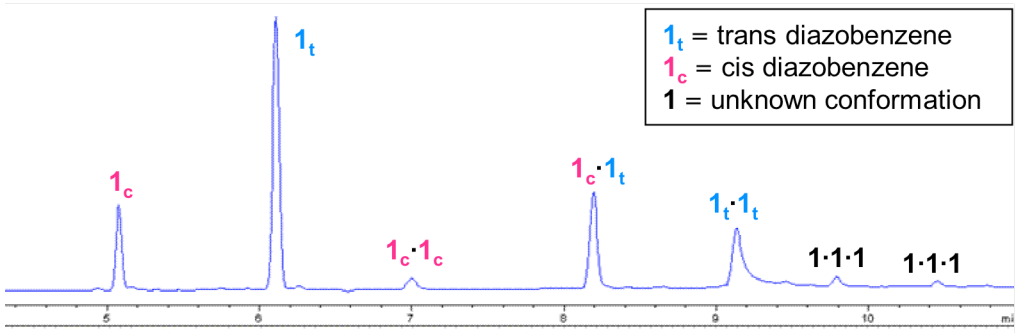



**Figure 3.4.** HPLC traces at 280 nm of a DCL of **1** at equilibrium.

At lower monomer concentrations (1 mM) the library generated is primarily composed of monomeric species. Each library member was observed in both the cis and trans conformations and the isomers were easily identified due to different retention times and a large difference in absorbance of trans and *cis*-azobenzene at 320 nm (Figure 3.5).[24] The cyclic hydrazones were found in a cis to trans ratio of about 20:80 at thermodynamic equilibrium. To ascertain the reversible nature of the library and confirm that the reaction mixture had reached equilibrium, the library was also generated from pure cyclic trans monomer. As expected, the final product distribution was the same.

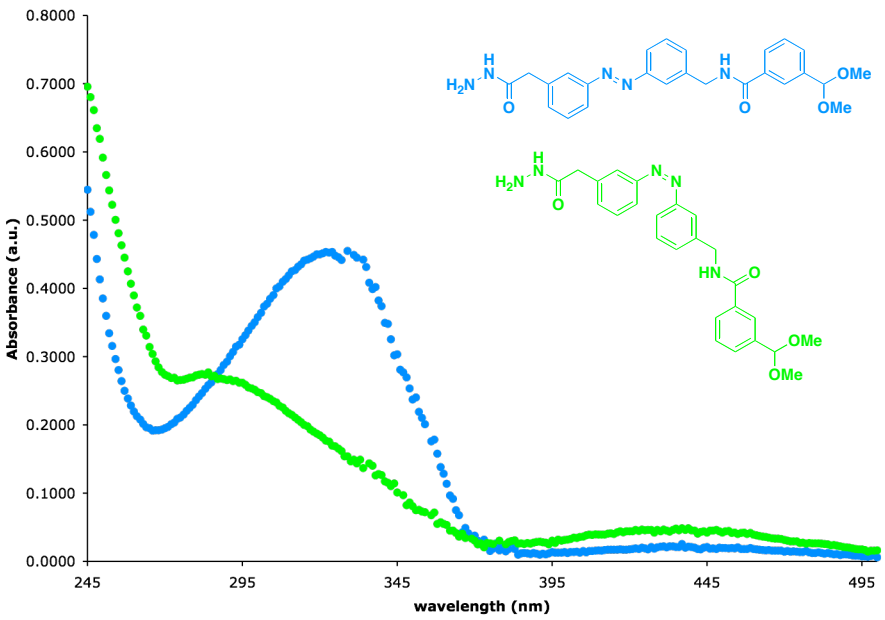



**Figure 3.5.** UV-Vis spectra of monomer **1** (0.022 μM) in the thermal state (approximately 80:20 *(E)*:*(Z)*; blue) and in the photostationary state after irradiation at 365 nm (approximately 15:85 *(E)*:*(Z)*; green). Spectra were taken in $CHCl_3$ at room temperature with a scan speed of 120 nm/min.

**iii. Library diversification with proline based hydrazine monomers.** To increase the structural diversity of the library and expand upon the number of host macrocycles generated, we chose to synthesize and investigate other hydrazine functionalized monomers, specifically proline based monomers **2** and **3** (Figure 3.6).[25] The rigid proline based building blocks **2** and **3** differ only in the position of the dimethoxyacetal substituent (meta and para, respectively). The necessary functional groups for hydrazone exchange were similarly installed in these monomers, following modified literature procedures.[19,25] The flexible azobenzene building block **1** was then reacted with either monomer **2** or **3**, as the mass degeneracy of monomers **2** and **3** prevented them from being combined in single DCL.

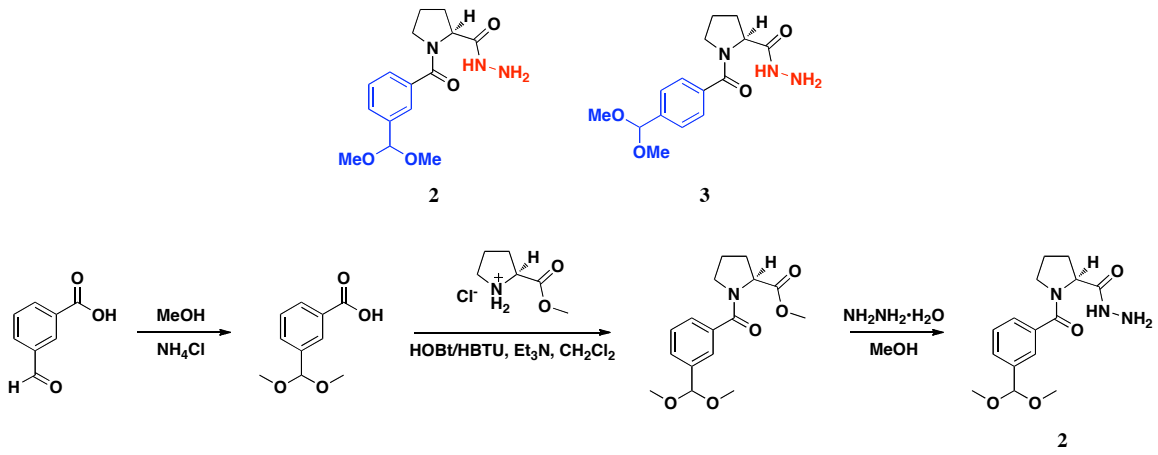



**Figure 3.6.** Proline building blocks **2** and **3** functionalized with a hydrazide (red) and protected aldehyde (blue), and the synthesis of **2**.

The DCLs were prepared (4 mM in each monomer) in a $CHCl_3$-DMSO (85:15 v/v) solution, as the proline monomer was found to assist with solubility and $CHCl_3$ is a more amenable solvent for further templation studies. After equilibration, these multi-building block DCLs generated complex library distributions with upward of twenty identifiable species and up to pentameric cyclic oligomers in some cases (Figures 3.7 and 3.8).

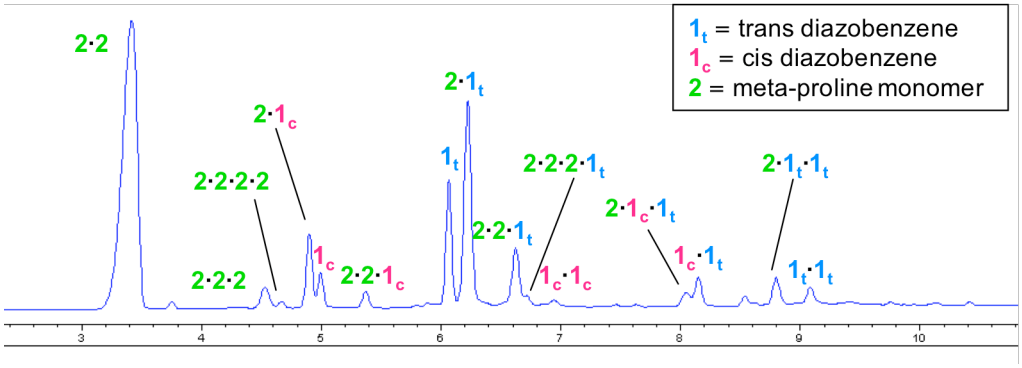



**Figure 3.7.** HPLC trace at 280 nm of a DCL of **1** and **2** at equilibrium.

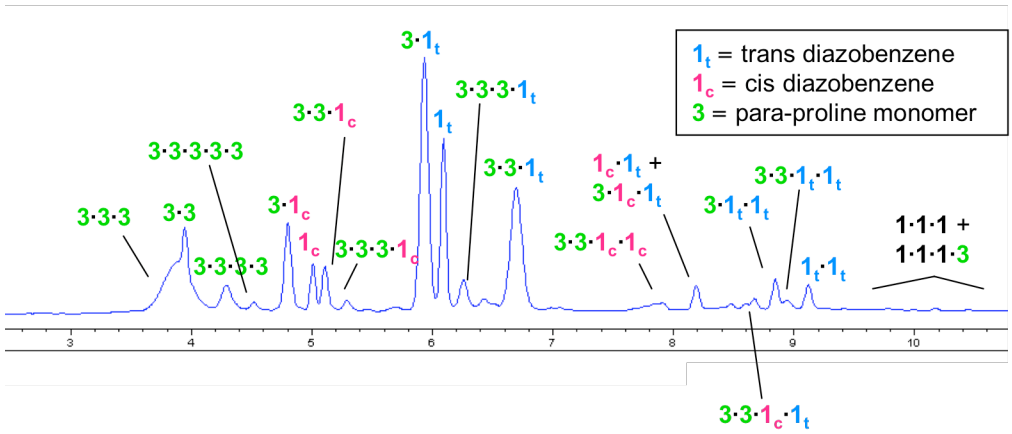



**Figure 3.8.** HPLC trace at 280 nm of a DCL of **1** and **2** (top) and **1** and **3** (bottom) at equilibrium.

The distinct differences between the distribution of species in the libraries generated with **2** versus **3** portray the sensitivity of DCLs to slight structural changes in the building blocks. Particularly in studying the fundamental aspects of molecular recognition, DCC can serve as an ideal system due to the changes in library composition that result from subtle variations in monomer structure. In this case, meta-substituted proline monomer **2** is found to self-sort to a much larger extent than **3**, giving rise to the proline homo-dimer (**2·2**) as the dominant library member in DCLs generated from **1** and **2**, whereas the hetero-dimer (**1·3**) is found to be the major species in DCLs generated from **1** and **3**.

**iv. Effect of photoisomerization on library distribution.** To investigate the dual nature of our libraries and their application in the development of photoswitchable hosts, we examined the effect of light-induced isomerization on the composition of our libraries. Upon irradiation of equilibrated DCLs at 365 nm for 5 minutes, each library distribution is shifted significantly in favor of the cis isomers. This assisted in confirming our assignments of the different isomers. Further isomerization does not occur with longer light exposure. After irradiation of the single building block library,

the azobenzene cyclic monomer species is present in a cis:trans ratio of 87:13 as observed by LC-MS (Figure 3.9). In this library a similar large increase of the cis/cis azobenzene homo-dimer ($\mathbf{1_c \cdot 1_c}$) is observed, accompanied by a significant decrease of the trans/trans azobenzene homo-dimer ($\mathbf{1_t \cdot 1_t}$) and a minor decrease in the cis/trans hetero-dimer ($\mathbf{1_c \cdot 1_t}$). Comparable changes are also observed with the trimer macrocycles (**1·1·1**), as well as with oligomers containing one or more proline monomer units in the multi-building block libraries (Figures 3.10 and 3.11).

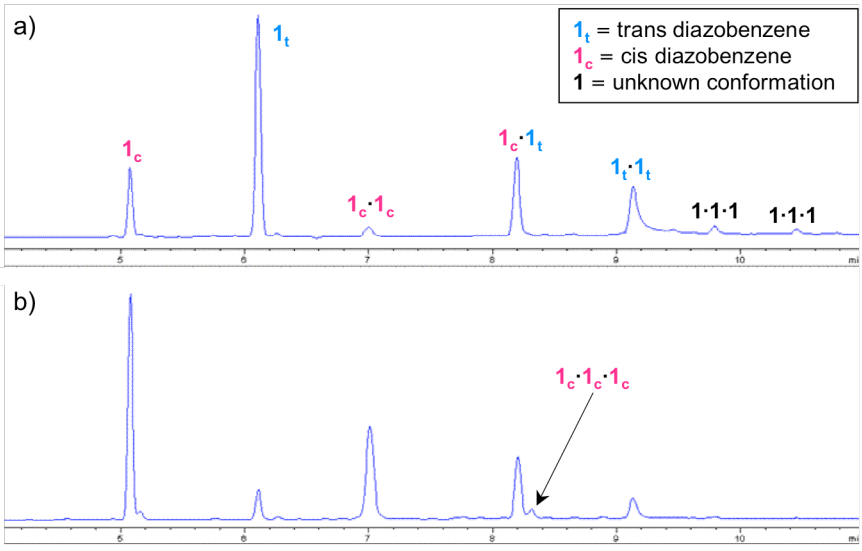



**Figure 3.9.** HPLC traces at 280 nm of a DCL of **1** (a) at equilibrium and (b) after isomerization. The chromatogram y-axes are on the same scale.

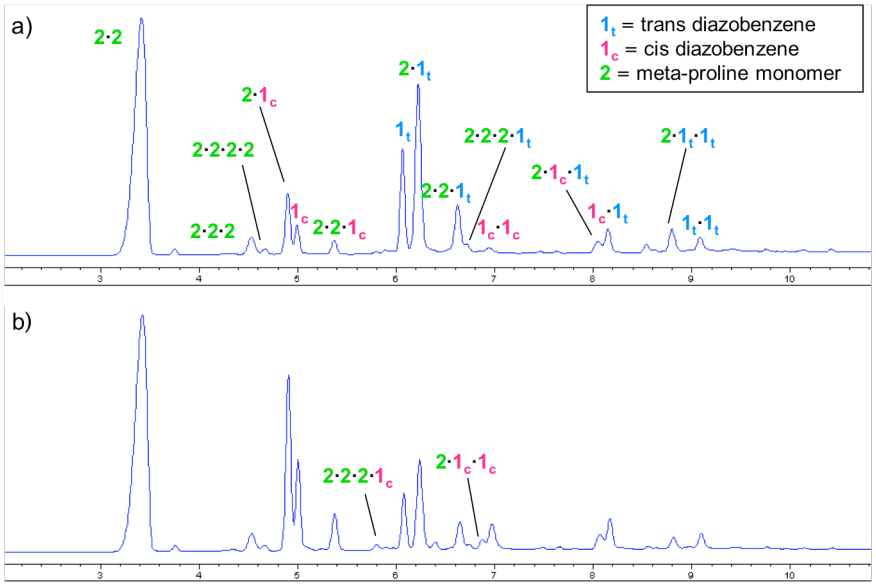



**Figure 3.10.** HPLC traces at 280 nm of a DCL of **1** and **2** (a) at equilibrium and (b) immediately after photoisomerization. The chromatogram y-axes are on the same scale.

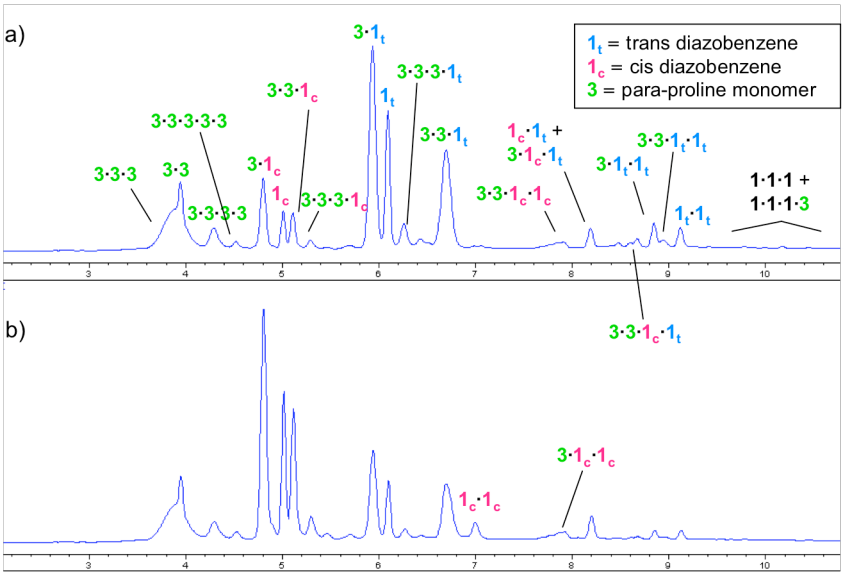



**Figure 3.11.** HPLC traces at 280 nm of a DCL of **1** and **3** (a) at equilibrium and (b) immediately after photoisomerization. The chromatogram y-axes are on the same scale.

Immediately after photoisomerization the libraries are stored in the dark, although with time they return to thermal equilibrium. The rate of conversion back to thermal

equilibrium seems to be solvent dependent, taking on the order of a week or less for the libraries generated in mostly $CHCl_3$, but longer for those in DMSO only. Hydrazone exchange within the irradiated DCL (Figure 3.12a) and relaxation back to thermal equilibrium are competitive processes, occurring simultaneously, and resulting in a DCL dominated by trans macrocycles (Figure 3.12b). In contrast, libraries which are equilibrated under photochemical conditions with repetitive irradiation allow for hydrazone exchange to occur in the absence of thermal relaxation (Figure 3.12c). By comparing parts a) and c) of Figure 3.12, both in the photostationary state, it is evident that the order of equilibration and photoisomerization along with the conditions under which the DCLs are equilibrated considerably influence the distribution of macrocycles. The integration of these two dynamic reversible processes permits various distributions of species to be generated under different conditions within a single DCL. Such versatility within a library may prove to be valuable in expanding the applications of DCC, as well as in the identification of photoswitchable receptors in this context.

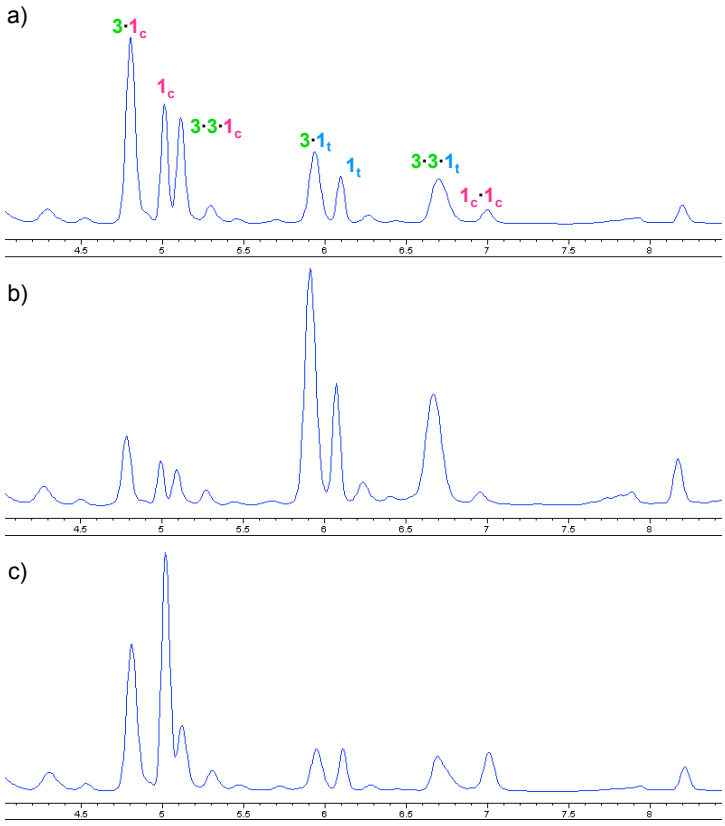



**Figure 3.12.** Part of the HPLC traces at 254 nm of a DCL of **1** and **3** (a) immediately after irradiation of an equilibrated library (b) six days after irradiation where there has been adequate time for thermal relaxation, and (c) equilibrated in the photostationary state with repetitive isomerization. The chromatogram y-axes are on the same scale.

Azobenzene monomeric building blocks which were subjected to photoisomerization before cyclization were also investigated. As expected, upon acid catalysis the irradiated monomers underwent simultaneous cyclization and thermal relaxation, which over time resulted in DCLs identical to those generated with the thermodynamically favored trans monomers.

**v. Templation studies with a polyproline peptide.** Because photoisomerization converts the library to a photostationary state rather than a thermally equilibrated state, if a guest binds a specific cis macrocycle, it would be expected to inhibit conversion of that

receptor back to its trans conformation due to favorable binding interactions (Figure 3.13). Although this is not amplification in the traditional sense, it would nonetheless result in an increased amount of the cis receptor in the templated library versus the corresponding untemplated DCL. This is indeed what was observed when the libraries were equilibrated in the presence of a pentaproline guest.

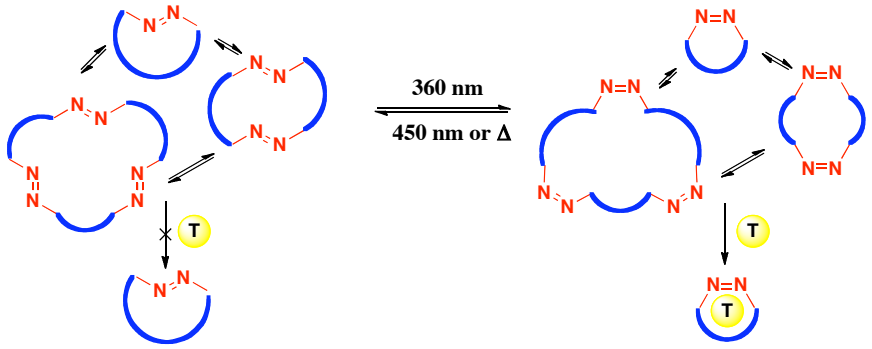



**Figure 3.13.** Illustration of the conformation of an azobenzene receptor controlling binding to a specific template, resulting in an observed stabilization of the cis conformer via binding.

We chose to investigate an oligoproline peptide guest due to the role of polyproline helices in many important protein-protein interactions.[26] A photoswitchable DCL consisting of monomers **1** and **2** was thermally equilibrated in the presence of a five-residue oligoproline peptide (Figure 3.14).

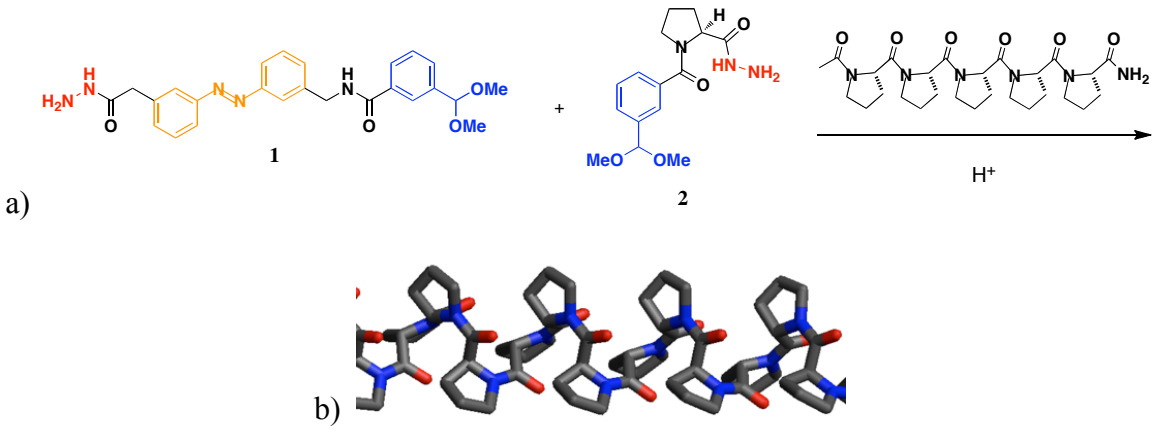



**Figure 3.14.** (a) Templation with a 5-residue polyproline peptide. (b) Side view of a polyproline type I right-handed helix, generally the more common conformation in organic solvents.[27] However, we have no direct evidence that the polyproline in our system is taking on a type I turn.

At thermal equilibrium, the $\mathbf{1_c \cdot 1_t}$ heterodimer is found to be amplified by about 46% relative to the untemplated library (Figure 3.15a and 3.15b). Immediately following photoisomerization of the equilibrated library, a small amplification of both $\mathbf{1_c \cdot 1_t}$ and $\mathbf{1_c \cdot 1_c}$ is observed relative to the untemplated library, yet there has been little time for re-equilibration of the newly isomerized macrocycles at this point (Figure 3.15c and 3.15d). In returning to a state of thermal equilibrium, a slower rate of thermal relaxation of some library members was observed in comparison to the untemplated library, indicative of favorable binding interactions with the polyproline guest. This rate difference is most apparent with the cyclic cis monomer $\mathbf{1_c}$. As seen in Figures 3.15e and 3.15f along with Figure 3.16, three days after the DCLs were irradiated at 365 nm, host $\mathbf{1_c}$ is significantly amplified in the polyproline templated DCL. As expected, this amplification is observed in DCLs containing monomer **1** and either proline monomer (**2** or **3**). In contrast, $\mathbf{1_t}$ is not amplified in the thermally equilibrated library, suggesting that the conformation of the azobenzene controls binding. The fact that $\mathbf{1_c}$ is only amplified after

photoisomerization is likely a result of not only the low concentration of *cis*-azobenzene macrocycles at thermal equilibrium, but also the competitive equilibria at play. Presumably at higher concentrations of the cis monomer, the favorable interactions which result in templation of $\mathbf{1_c}$ out-compete other equilibria. This type of phenomenon is not unprecedented in dynamic combinatorial libraries,[28] and reinforces the fact that the thermally equilibrated and photoisomerized libraries represent different libraries with different behaviors which indeed should be considered independently.

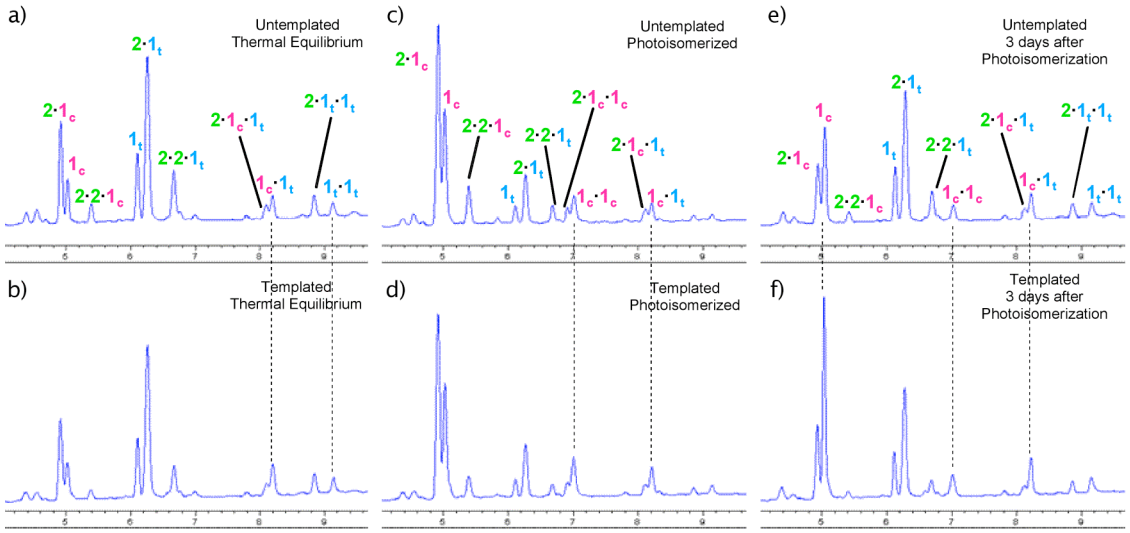



**Figure 3.15.** Part of the HPLC trace at 254 nm of a DCL made from **1** and **2** (a) at thermal equilibrium without an added template (b) at thermal equilibrium in the presence of a pentaproline peptide (25 mM) (c) immediately after irradiation, untemplated, (d) immediately after irradiation with a pentaproline peptide, (e) three days after irradiation, untemplated, and (f) three days after irradiation with a pentaproline peptide. Note that figures (e) and (f) are not meant to represent DCLs which have returned to thermal equilibrium. Amplified species are indicated with the dashed lines between untemplated and templated libraries. The chromatogram y-axes are on the same scale.

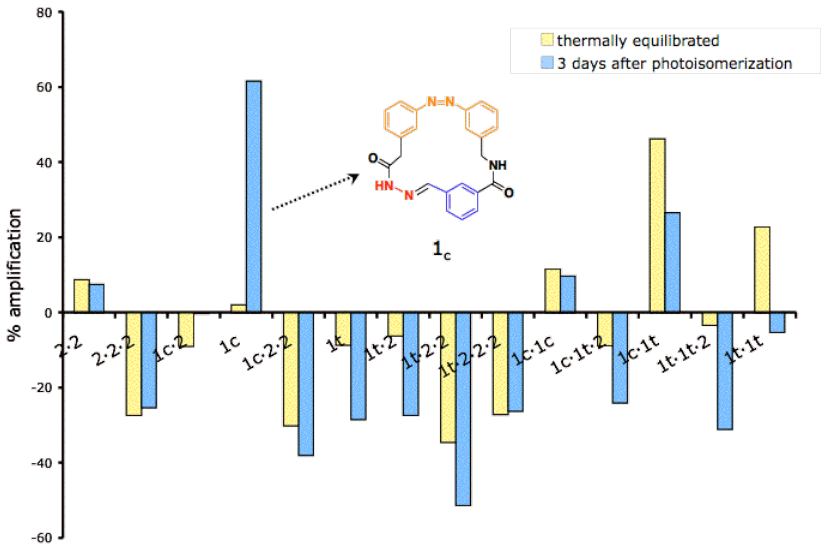



**Figure 3.16.** Comparison between the extent of amplification of each library member in the polyproline templated library (made from **1** and **2**) at thermal equilibrium (yellow) and 3 days after irradiation (blue). (Percent amplification = [(% area of library member in templated DCL - % area of library member in untemplated DCL) / % area of library member in untemplated DCL] * 100).

Unfortunately, as is often the case with hydrazone libraries, the binding constants of the isolated host-guest complex could not be measured because the host is not soluble under non-equilibrating (ie non-acidic) conditions, except in DMSO, which disrupts binding.[29]

The observed stabilization of a *cis*-azobenzene via binding is not unprecedented. Such stabilization has been previously reported in the context of binding tethered, photoswitchable maleimide-azobenzene-glutamate ligands.[30] Polyproline peptides and other such guests which form more favorable binding interactions with one isomer of a specific host are especially attractive due to the potential to photomodulate the binding to these receptors, therefore achieving a larger degree of control over such molecular recognition processes.

**vi. Other templates investigated and limitations of azobenzene hydrazone DCLs.**

In our search for a template that would perturb the equilibrium, resulting in the stabilization and amplification of a specific receptor due to favorable binding interactions, a variety of templates were explored, some of which are shown in Figure 3.17. Cationic aromatic ammonium groups were investigated due to their potential to interact favorably with the designed monomers via interactions such as aromatic-aromatic or cation-π interactions, a few of which have already been shown by Sanders and coworkers to bind macrocycles composed of proline monomer **2**.[25] In addition, various biologically relevant molecules in addition to the polyproline peptide were considered.

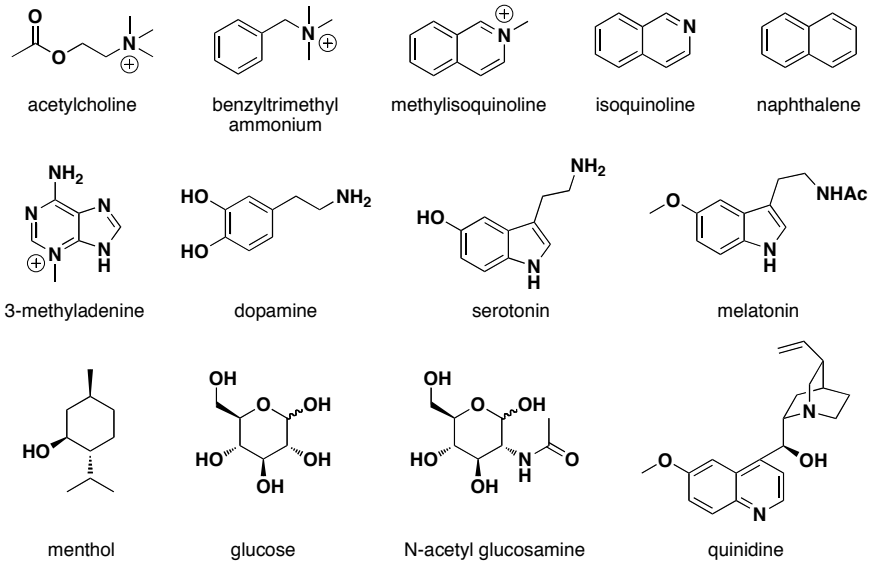



**Figure 3.17.** Other template molecules investigated in azobenzene hydrazone DCLs.

In our screens of different guests, there were various factors which limited those templates that were compatible with the designed libraries. First, solubility was a major factor, as many biological molecules in particular were not soluble under the conditions for equilibration. Limited solubility was met with both glucose and N-acetyl glucosamine, as well as hormones dopamine and serotonin. Other guests were soluble

initially but led to precipitation over time during library equilibration, whereby an accurate comparison to the untemplated reaction was no longer possible. Some amines such as 3-methyl adenine became soluble only upon protonation with the addition of TFA.

In addition, chromatographic separation of the template from the rest of the library is required. Co-elution of the template and library macrocycles prevents analysis of the library composition, as the template is present in excess and its absorption generally exceeds that of the library members. This was problematic with hydrophobic guests such as naphthalene, whereas more polar guests tended to have a much earlier retention time. Furthermore, guests containing functional groups that could participate in hydrazone exchange had to be avoided, while photolytic stability of the template was also required. Both hormones melatonin and dopamine are light sensitive, preventing analysis of such templated libraries upon photolysis.

The azobenzene moiety can undergo surprisingly facile reduction to hydrazobenzene in the presence of reducing agents, further limiting the scope of templates that could be explored.[31] This facile reduction of azobenzene to hydrazobenzene was observed in various instances when the library was inadvertently exposed to a reducing agent at some point during the LC-MS analysis. This also makes azobenzene isomerization incompatible with the commonly used reversible reactions, disulfide and thioester exchange.

The results of templation studies with other guests, particularly methylisoquinoline and benzyltrimethylammonium, although not entirely understood, are also worth noting in providing a thorough discussion of the template limitations with these azobenzene

libraries. In a DCL with monomers **1** and **2** in the presence of benzyltrimethylammonium chloride, significant amplification of $\mathbf{2_3}$ and $\mathbf{1_t \cdot 2}$ was observed (Figure 3.18). Amplification of $\mathbf{2_3}$ was anticipated based on literature precedence.

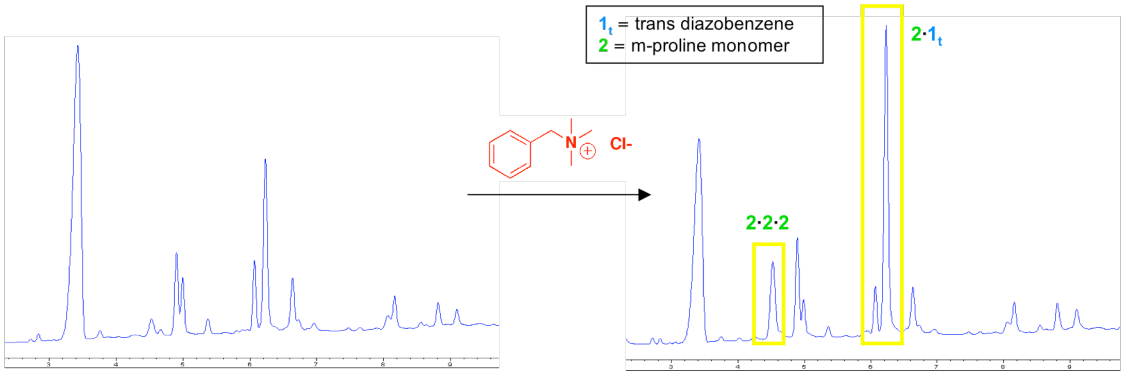



**Figure 3.18.** HPLC traces at 280 nm of an untemplated DCL of **1** and **2** (left) and in the presence of 100 mM benzyltrimethylammonium chloride (right) at thermal equilibrium. The chromatogram *y*-axes are on the same scale.

In addition to the template perturbing the library at thermal equilibrium, after photolysis, the **1·2** *cis*-heterodimer converts back to the thermodynamically favored trans conformation faster than in the untemplated library, reinforcing that there may be a favorable binding interaction with the *trans*-heterodimer (Figure 3.19).

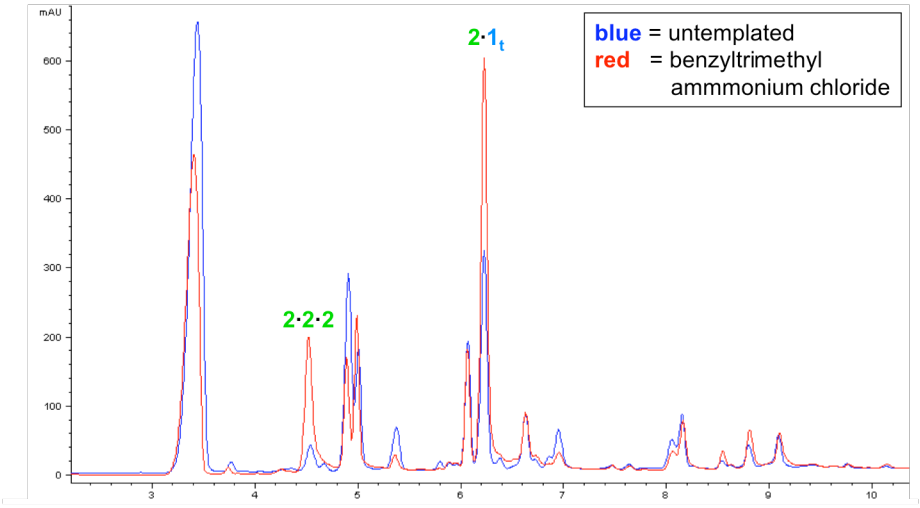



**Figure 3.19.** HPLC traces at 280 nm of an untemplated DCL of **1** and **2** (blue) and in the presence of 100 mM benzyltrimethylammonium chloride (red) 1 day after photolysis.

Comparable results were also observed with libraries containing monomers **1** and **3**, where the *trans*-heterodimer **1·3** was amplified at thermodynamic equilibrium, and after photolysis underwent a faster rate of thermal relaxation than in the untemplated library as well. It was proposed that the position of the acetal in the proline monomer did not invoke specificity in binding benzyltrimethylammonium. Acetylcholine was also investigated with monomers **1** and **3**, and similar amplification was observed to that of benzyltrimethylammonium, although to a somewhat lesser extent.

In the case of templation with methylisoquinoline, the **1·2** *trans*-heterodimer again appeared to be amplified at thermal equilibrium. However, in this case none of the macrocycles containing *cis*-**1** were observed (Figure 3.20). Upon photolysis, there were no distinct changes in the methylisoquinoline templated library, and the expected trans to cis isomerization was not observed. Suspect that the methylisoquinoline may have been absorbing the 365 nm light, a control experiment was performed where the sequence of events was reversed, and methylisoquinoline was added *after* isomerization as opposed to prior experiments where photolysis was performed in the presence of guest. The end result remained unchanged, as no cis macrocycles were observed after addition of methylisoquinoline.

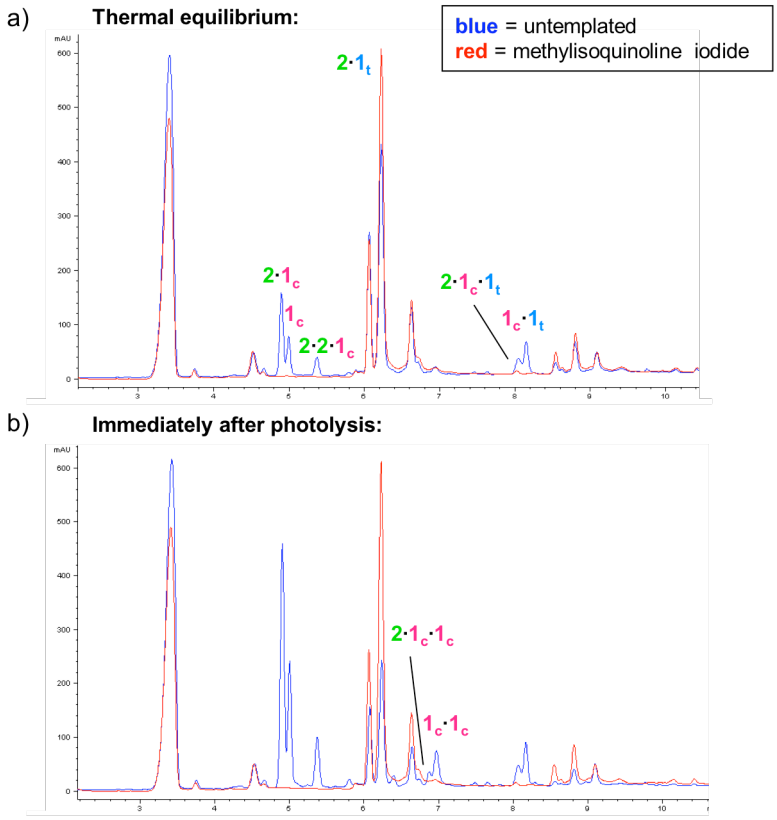
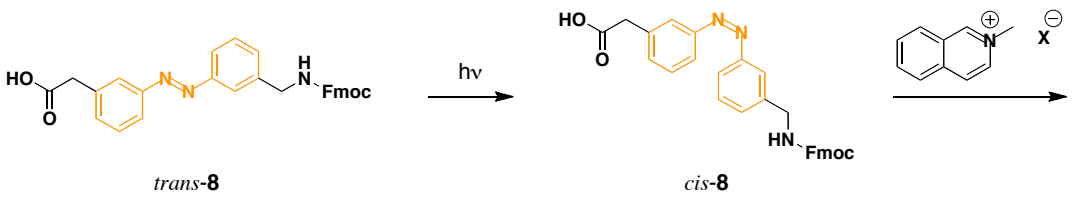



**Figure 3.20.** HPLC traces at 280 nm of a DCL of **1** and **2** untemplated and in the presence of 40 mM methylisoquinoline iodide (a) at thermal equilbrium and (b) immediately after photolysis.

To further understand this result and confirm that the amplification was indeed due to methylisoquinoline binding, hydrazide monomer **1** was replaced with the corresponding Fmoc protected amino acid **8**, which lacks the functionality to undergo hydrazone exchange and therefore form cyclic macrocycles. Amino acid **8** was prepared simply by subjecting precursor **4** to trifluoroacetic acid. It was anticipated that the cis:trans ratio of **8** would not be affected by methylisoquinoline.




*trans*-**8** *cis*-**8**

**Figure 3.21.** Control experiments with azobenzene amino acid **8** (X = iodide or triflate counterion).

Solutions of **8** were prepared in 15:85 DMSO:$CHCl_3$ and subjected to photolysis, after which methylisoquinoline iodide was added (Figure 3.21). Prior to the addition of TFA, the amino acid in the absence and presence of methylisoquinoline was photolyzed as expected, converting to 85% and 83% *cis*-azobenzene respectively immediately after photolysis (as determined by UV integration at 280 nm). The reactions were allowed to thermally relax in the dark, and the rate of thermal relaxation appeared to be the about the same in both cases. After eleven days the ratio of *cis*-**8**:*trans*-**8** had converted to about 45:55. In order to more directly mimic the equilibration conditions used for exchange, 100 mM of TFA was added to both solutions. Immediate LC-MS analysis revealed that while the solution without methylisoquinoline remained at a *cis*-**8**:*trans*-**8** ratio of 45:55, in contrast, the solution with methylisoquinoline had been converted entirely to *trans*-**8**.

To explore the dependence on the nucleophilic iodide counterion, methylisoquinoline was prepared with a non-nucleophilic triflate counterion and the control experiment was repeated (with TFA added from the beginning). Similarly, no *cis*-**8** was observed in the presence of methylisoquinoline triflate, indicating that the results appear to be independent of the counterion. In contrast, the same control experiment was performed in the presence of tetrabutylammonium with either an iodide, bromide, or fluoride counterion. In this case the rate of thermal relaxation of *cis*-**8** did seem to differ in the

presence of different counterions, with the general trend of $N(Bu)_4^+I^- > N(Bu)_4^+Br^- > N(Bu)_4^+F^-$. The isomerization occurs most rapidly with the nucleophilic iodide counterion, whereas the rate of relaxation in the presence of $N(Bu)_4^+F^-$ is close to that of the solution of **8** by itself, suggesting the presence of a nucleophile may indeed be important in this acid promoted isomerization. It is evident that the apparent amplification in the presence of methylisoquinoline is not the result of binding to the azobenzene receptor macrocycles (particularly $\mathbf{1_t \cdot 2}$), but instead a TFA promoted mechanism by which methylisoquinoline facilitates the conversion of all *cis*-azobenzene molecules to their corresponding trans isomers.

The same control experiment was performed with amino acid **8** and benzyltrimethylammonium chloride, which also showed promising amplification of receptor $\mathbf{1_t \cdot 2}$, both at thermal equilibrium and during thermal relaxation. The solutions were analyzed daily by LC-MS, and although less drastic than with methylisoquinoline, there was a noticeably faster rate of thermal relaxation from *cis*-**8** to *trans*-**8** in the presence of benzyltrimethylammonium (Figure 3.22). In the presence of certain cationic guests and their corresponding counterions, it is clear that *cis*-azobenzene isomers are rapidly converted to their trans counterparts under the acidic conditions necessary for hydrazone exchange. This clearly perturbs the dynamic equilibria in our libraries, yet is unrelated to favorable binding interactions, and limits the range of guests that can be studied.

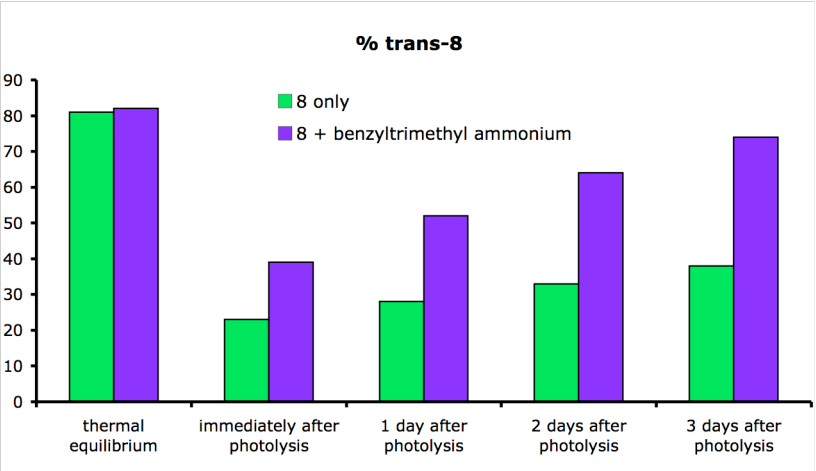



**Figure 3.22.** Amount of *trans*-**8** present in solution at different time points during thermal relaxation, by itself (green) and with benzyltrimethyl ammonium (purple).

## C. Conclusions

This work reports the design of an azobenzene based hydrazide monomer that can undergo thermodynamically controlled oligomerization and cyclization under acidic conditions, either in single or multi-building block libraries. Although the thermally equilibrated DCLs favor the trans isomers, we have shown that we can photochemically control the library distribution by irradiating the systems at appropriate wavelengths. The incorporation of these two reversible processes in a single library is advantageous in that it offers a higher degree of control over the library composition and its organization in the investigation of potential targets. In exploring the templation of a polyproline peptide, we have demonstrated that we can stabilize and amplify an inherently less stable cis azobenzene macrocycle through templation with an appropriate guest. The development of hosts for such proline rich sequences via DCC may help to elucidate their role as potential protein interaction domains.

Photoresponsive libraries of this type are useful in that, despite the nature of the guest, binding can be controlled by irradiation at appropriate wavelengths. This shows the potential to, in one system, both identify and synthesize novel host receptors and take advantage of the optical properties of azobenzene to control binding.

## D. Experimental Section

**i. Synthesis of azobenzene monomer.** *(E)-methyl 2-(3-((3-((((9H-fluoren-9-yl)methoxy)carbonylamino)methyl)phenyl)diazenyl)phenyl)acetate (**5**).* In 27 mL of methanol containing 0.1M $H_2SO_4$, **4** (0.29 g, 0.53 mmol) was dissolved and the solution was refluxed for six hours. After removal of the solvent, water was added to precipitate the product and it was extracted three times with EtOAc. The organic extracts were dried over $MgSO_4$ and the solvent was evaporated to obtain an orange solid (94%). $^1$H NMR ($CDCl_3$, 300 MHz): δ 7.83-7.81 (m, 4H), 7.75 (d, $J$ = 7.5 Hz, 2H), 7.59 (d, $J$ = 7.5 Hz, 2H), 7.50-7.44 (m, 2H), 7.41-7.36 (m, 4H), 7.31-7.26 (m, 2H), 5.26 (t, $J$ = 6.0 Hz, 1H), 4.47- 4.45 (m, 4H), 4.23 (t, $J$ = 6.8 Hz, 1H), 3.72 (s, 2H), 3.70 (s, 3H). HRMS-ESI(+) m/z $[M+H]^+$ calculated for $C_{31}H_{27}N_3O_4$ = 506.2080, found = 506.2062; m/z $[M+Na]^+$ calculated for $C_{31}H_{27}N_3NaO_4$ = 528.1899, found = 528.1907.

*(E)-methyl 2-(3-((3-(aminomethyl)phenyl)diazenyl)phenyl)acetate (**6**).* To a solution of **5** (0.26 g, 0.52 mmol) in 5.2 mL dichloromethane, tris(2-aminoethyl)amine (3.9 mL, 26 mol) was added. The mixture was stirred for one hour and then washed with saturated NaCl followed by phosphate buffer pH 5.5. The organic extracts were dried over $MgSO_4$. After the solvent was evaporated, the resulting oil was purified by flash chromatography (MeOH with 10% $NH_4OH/CH_2Cl_2$ 1:9) to obtain 0.18 g (68%) of a

bright orange oil. The product was isolated as a mixture cis and trans isomers, with the cis isomer ranging from 10 to 20% as determined by NMR integrations. $^1$H NMR ($CDCl_3$, 300 MHz): δ 7.85-7.77 (m, 4H), 7.49-7.37 (m, 4H), 3.96 (s, 2H), 3.72 (s, 2H), 3.70 (s, 3H). HRMS-ESI(+) $[M+H]^+$ m/z calculated for $C_{16}H_{17}N_3O_2$ = 284.1399, found = 284.1402.

*(E)-methyl 2-(3-((3-((3-(dimethoxymethyl)benzamido)methyl)phenyl)diazenyl)phenyl) acetate (**7**)*. A solution of 0.022 g (0.112 mmol) of 3-carboxybenzaldehyde dimethoxyacetal, 0.015 g (0.112 mmol) of HOBt, and 0.042 g (0.112 mmol) of HBTU in 700 μL of DMF was added to **6** (0.032 g, 0.112 mmol). To the mixture was added 47 μL (0.336 mmol) of triethylamine, and the solution was stirred for four hours. After the addition of brine, the product was extracted with EtOAc. The organic phase was washed with sodium bicarbonate, dried over $MgSO_4$, and the solvent was evaporated. The resulting orange oil was purified by flash chromatography (MeOH with 10% $NH_4OH/CH_2Cl_2$ 1:19) to obtain 0.031 g (60%) of a bright orange oil. The product was again isolated as a similar mixture if isomers, with a small amount of the corresponding deprotected aldehyde (5-10%) which was carried through without a problem. $^1$H NMR ($CDCl_3$, 300 MHz): δ 7.86-7.80 (m, 6H), 7.57 (d, *J* = 7.5 Hz, 1H), 7.48-7.37 (m, 5H), 5.39 (s, 1H), 4.72 (d, *J* = 5.7 Hz, 2H), 3.71 (s, 2H), 3.69 (s, 3H), 3.29 (s, 6H). HRMS-ESI(+) $[M+Na]^+$ m/z calculated for $C_{26}H_{27}N_3NaO_5$ = 484.1849, found = 484.1861.

*(E)-3-(dimethoxymethyl)-N-(3-((3-(2-hydrazinyl-2-oxoethyl)phenyl)diazenyl)benzyl) benzamide **((E)-1)***. A solution of **7** (0.032 g, 0.069 mmol) in anhydrous methanol (0.8 mL) was treated with hydrazine monohydrate (33.5 μL, 0.69 mmol). The reaction was left overnight under nitrogen before removal of the solvent under vacuum to give an

orange oil which was purified by flash chromatography (MeOH with 10% $NH_4OH/CH_2Cl_2$ gradient, 1:39 to 1:19 to 1:9) to afford the monomer **1** (65%). $^1$H NMR ($CDCl_3$, 300 MHz): δ 7.88-7.77 (m, 6H), 7.58 (d, *J* = 7.5 Hz, 1H), 7.50-7.36 (m, 5H), 5.39 (s, 1H), 4.74 (d, *J* = 5.7 Hz, 2H), 3.63 (s, 2H), 3.30 (s, 6H). HRMS-ESI(+) $[M+Na]^+$ m/z calculated for $C_{25}H_{27}N_5NaO_4$ = 484.1961, found = 484.1958.

*(Z)-3-(dimethoxymethyl)-N-(3-((3-(2-hydrazinyl-2-oxoethyl)phenyl)diazenyl)benzyl) benzamide **((Z)-1)***. ***(E)-1*** was photolyzed at 365 nm. $^1$H NMR ($CDCl_3$, 300 MHz): δ 7.81 (s, 1H), 7.73 (d, *J* = 7.8 Hz, 1H), 7.53 (d, *J* = 7.8 Hz, 1H), 7.38 (t, *J* = 7.8 Hz, 1H), 7.14-7.01 (m, 3H), 6.75-6.58 (m, 5H), 5.71 (d, *J* = 12 Hz, 2H, *$NH_2$*), 5.37 (s, 1H), 4.46 (d, *J* = 5.7 Hz, 2H), 3.34 (s, 2H), 3.28 (s, 6H).

**ii. Synthesis of proline monomers.** *(S) N-(3-Dimethoxymethyl-benzoyl)-proline carboxylic acid hydrazide **(2)***. Monomer **2** was prepared according to literature procedures,[20] except the amide bond formation was accomplished with HBTU/HOBt as the coupling reagent rather than EDC, as described below. A solution of 0.059 g (0.302 mmol) of 3-carboxybenzaldehyde dimethoxyacetal, 0.041 g (0.302 mmol) of HOBt, and 0.115 g (0.302 mmol) of HBTU in 1.5 mL of DMF was added to .050 g (0.302 mmol) of L-Proline methyl ester hydrochloride. To the mixture was added 126 μL (0.906 mmol) of triethylamine, and the solution was stirred for 1.5 hours. After the addition of brine, the product was extracted with EtOAc. The organic phase was washed with sodium bicarbonate, dried over $MgSO_4$, and the solvent was evaporated to give 0.088 g (94%) of a colorless oil. The hydrazinolysis reaction of the methyl ester was carried out as previously reported to afford **2**.

*(S) N-(4-Dimethoxymethyl-benzoyl)-proline carboxylic acid hydrazide **(3)***. The synthesis of (S) *N*-(4-Dimethoxymethyl-benzoyl)-proline carboxylic acid hydrazide **(3)** was carried out using 4-carboxybenzaldehyde dimethoxyacetal and the same conditions as those reported above for monomer **2**.

**iii. $^1$H NMR spectra.**

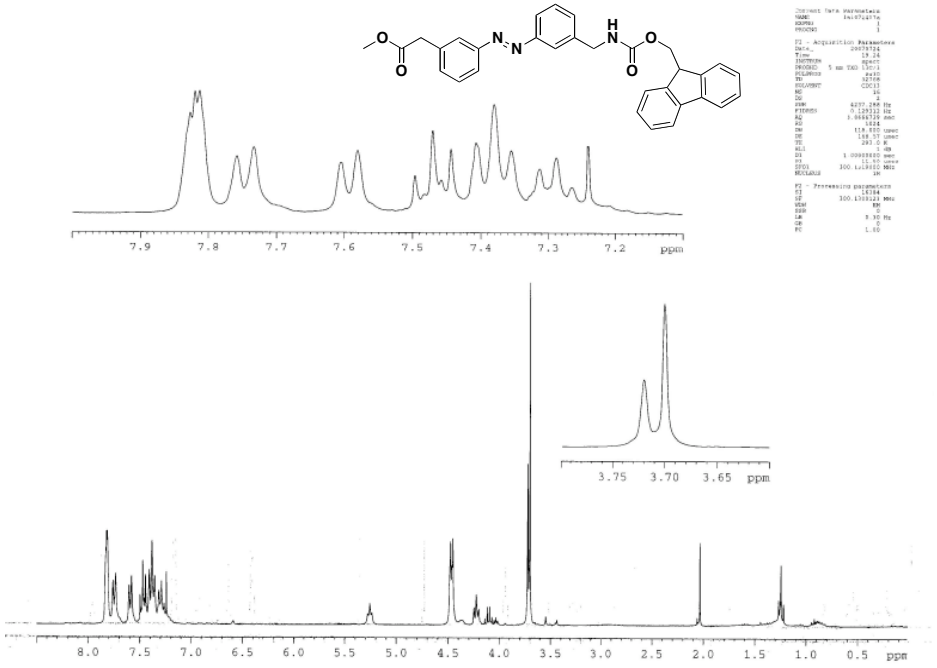



**Figure 3.23.** (*E*)-methyl 2-(3-((3-((((9H-fluoren-9-yl)methoxy)carbonylamino)methyl) phenyl)diazenyl)phenyl)acetate (**5**): $CDCl_3$, 300MHz.

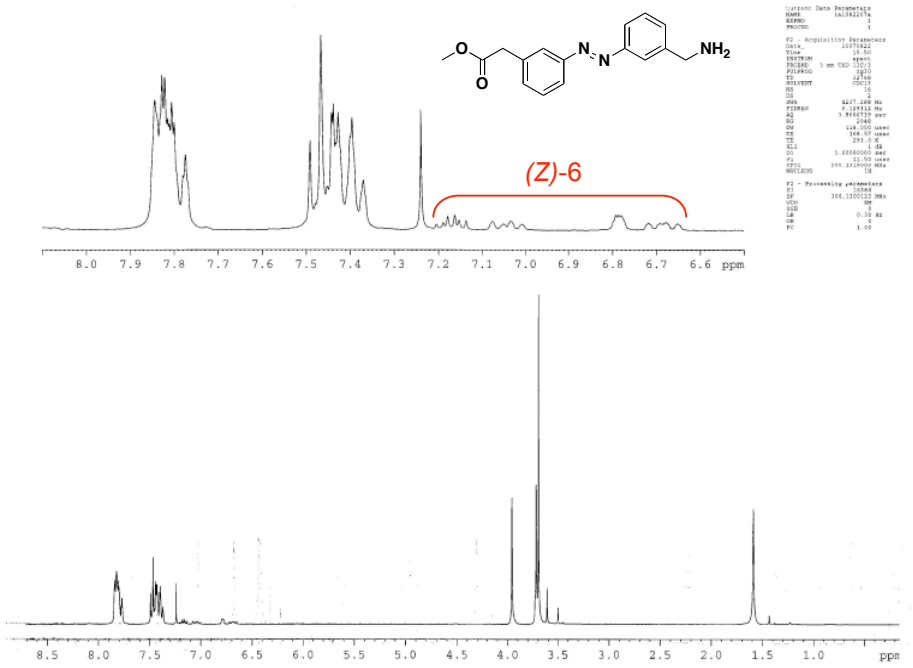



**Figure 3.24.** (*E*)-methyl 2-(3-((3-(aminomethyl)phenyl)diazenyl)phenyl)acetate (**6**): $CDCl_3$, 300MHz.

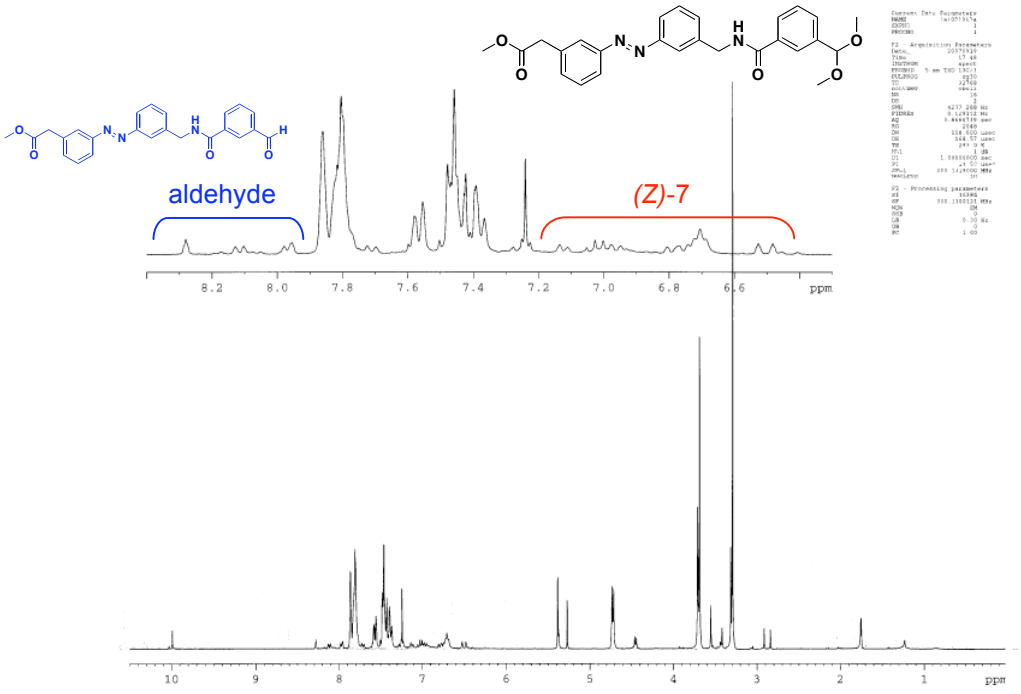



**Figure 3.25.** *(E)*-methyl 2-(3-((3-((3-(dimethoxymethyl)benzamido)methyl)phenyl) diazenyl)phenyl)acetate (**7**): $CDCl_3$, 300MHz.

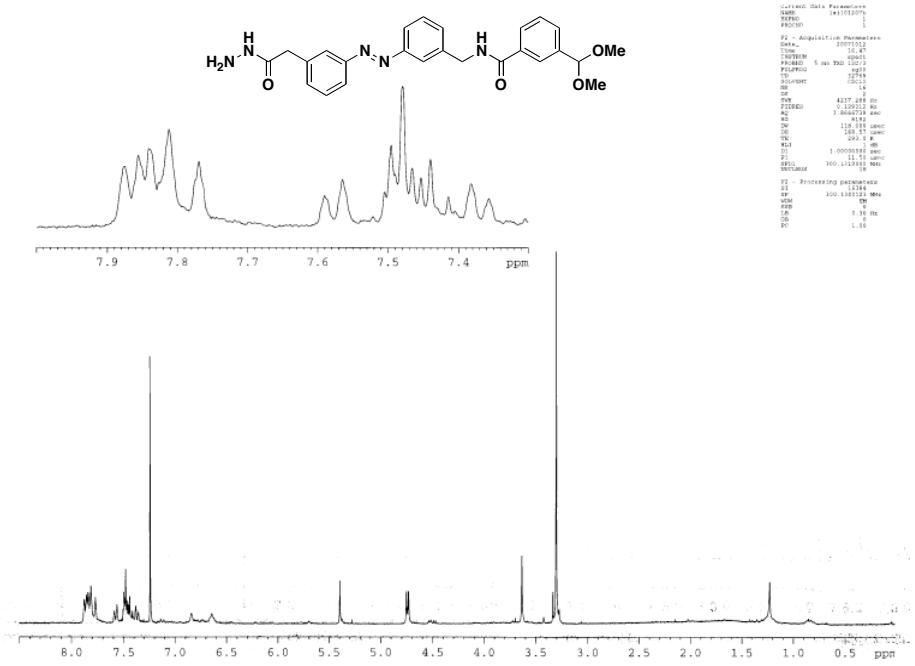


**Figure 3.26.** *(E)*-3-(dimethoxymethyl)-N-(3-((3-(2-hydrazinyl-2-oxoethyl)phenyl) diazenyl)benzyl) benzamide (*(E)*-**1**)): $CDCl_3$, 300MHz.

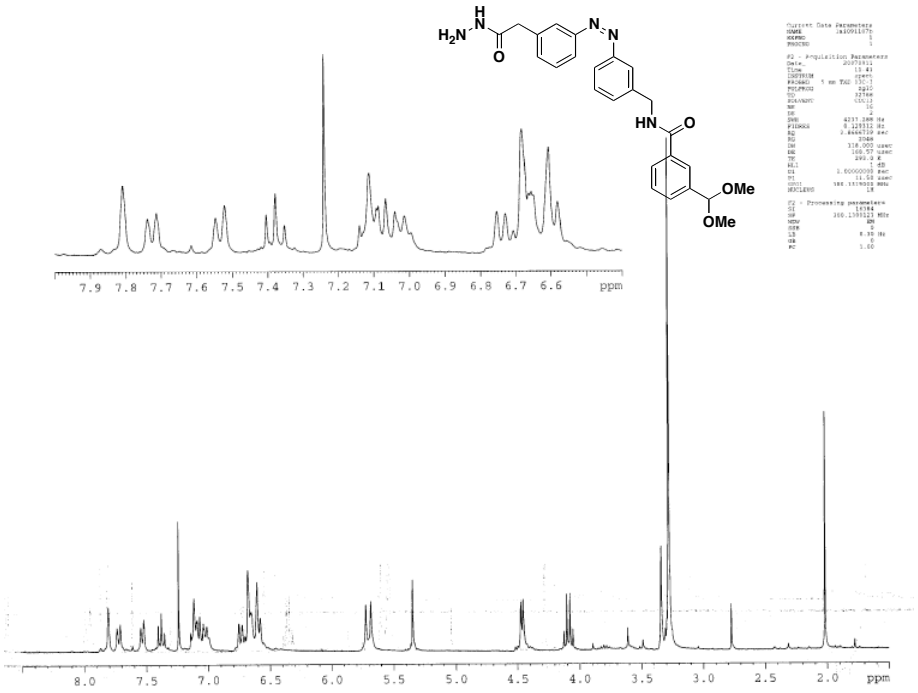


**Figure 3.27.** *(Z)*-3-(dimethoxymethyl)-N-(3-((3-(2-hydrazinyl-2-oxoethyl)phenyl) diazenyl)benzyl) benzamide (*(Z)*-**1**)): $CDCl_3$, 300MHz.

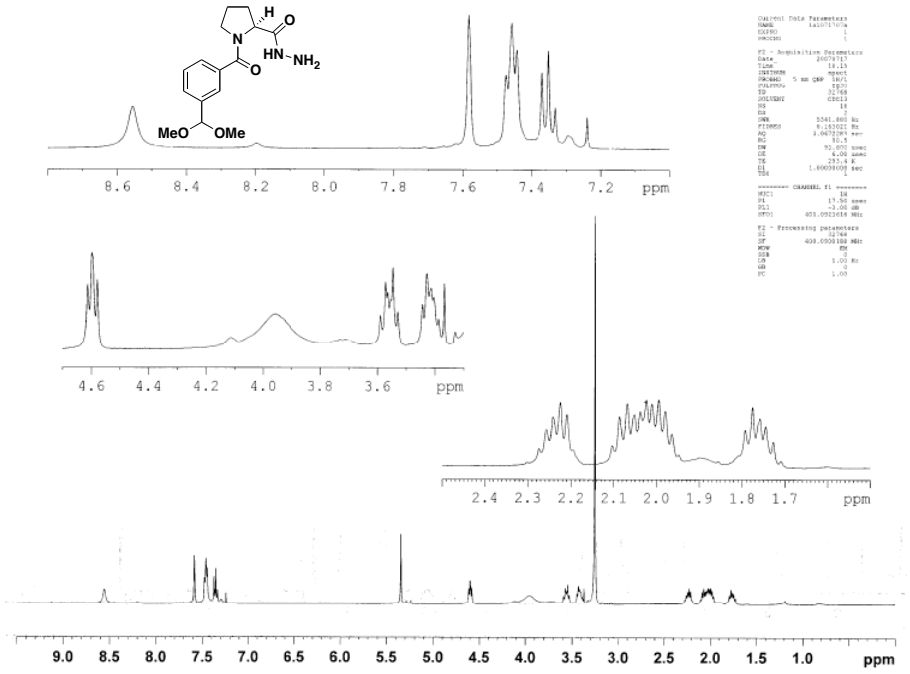



**Figure 3.28.** (*S*) *N*-(3-Dimethoxymethyl-benzoyl)-proline carboxylic acid hydrazide (**2**): $CDCl_3$, 400MHz.

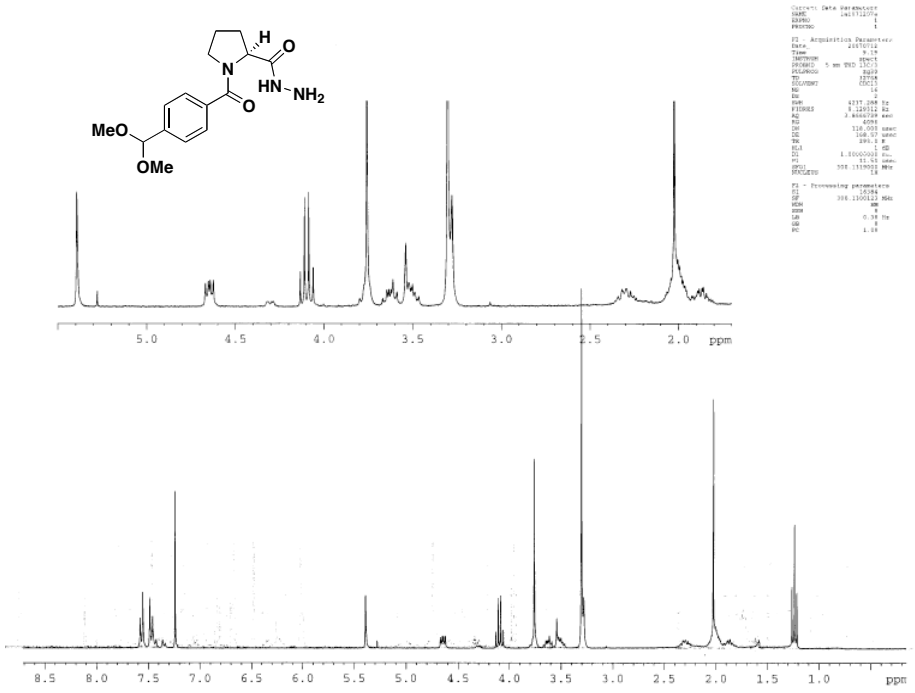



**Figure 3.29.** (*S*) *N*-(4-Dimethoxymethyl-benzoyl)-proline carboxylic acid hydrazide (**3**): $CDCl_3$, 300MHz.

**iv. Synthesis of polyproline peptide (Ac-Pro-Pro-Pro-Pro-Pro-$NH_2$).** The peptide synthesis was performed on a Tetras Peptide Synthesizer using Applied Biosystems PAL-PEG amide resin. The peptide was synthesized on a 0.0215 mmol scale (50 mg resin). Coupling reagents were HOBt/HBTU in DMF. The N-terminus was acylated with a solution of 5% acetic anhydride and 6% 2,6-lutidine in DMF. Cleavage was performed by hand with a cocktail of 95% TFA/2.5% triisopropylsilane/2.5% $H_2O$. The peptide was purified by semipreparative reverse phase HPLC on a C18 column at a flow rate of 4 mL/min. The purification was achieved using a linear gradient of A and B (A: 95% $H_2O$/5% $CH_3CN$ with 0.1% TFA; B: 95% $CH_3CN$/5% $H_2O$ with 0.1% TFA) and elution was monitored at 214 nm. Once purified, the peptide was lyophilized and characterized by ESI-MS. ESI-MS (+) m/z $[M+H]^+$ calculated for $C_{27}H_{40}N_6O_6$ = 545.30, found = 545.3.

**v. Synthesis of *N*-methylisoquinoline triflate.** To a flame dried flask purged with nitrogen, 300 μl (2.53 mmol) of isoquinoline and 2.5 mL of dry dichloromethane was added and the solution was cooled to 0°C. To a second flame dried flask purged with nitrogen, 2.5 mL of dry dichloromethane was added followed by careful addition of 286 μl (2.53 mmol) of methyltriflate. The methyltriflate solution was transferred via cannulation to the cold isoquinoline solution, and the reaction was stirred overnight under nitrogen, gradually warming up to room temperature. To the solution, 20 mL of ether was added, resulting in a precipitate which was filtered to obtain 0.704 g (95%) of a tan solid. The product appeared pure enough for most purposes, but can be recrystallized from acetonitrile/ethanol.

**vi. Dynamic Combinatorial Chemistry and LC-MS analysis. (a) DCL with monomer 1.** The single component library was prepared by making a 4 mM building block solution of **1** in DMSO containing 100 mM of TFA. The resulting solution equilibrated at room temperature for at least 4 days. The reaction was monitored daily by LC-MS (3 μL injections). Separations were performed using $H_2O$-acetonitrile gradients with 0.2% formic acid (t=0 min: 100% water, flow rate 1.0 mL/min; t=2 min: 75% water, flow rate 1.0 mL/min; t=3.5 min: 66% water, flow rate 1.0 mL/min; t=6.5 min: 57% water, flow rate 1.5 mL/min; t=11 min: 30% water, flow rate 1.5 mL/min), with the left column temperature set to 40°C and the right column temperature set to 50°C to optimize separation. Multiple wavelengths were used for analysis (220 nm, 254 nm, 280 nm, and 320 nm). The cis and trans isomers were assigned based on the strong absorbance of trans azobenzene at 320 nm, along with the shift in the distribution of isomers upon isomerization.

**(b) DCL with monomers 1 and 2 or 1 and 3.** The mixed library was prepared by making a 1:1 solution of the building blocks (4 mM each) in CHCl3-DMSO (85:15 v/v) containing 100 mM of TFA. The resulting solution was stirred at room temperature for at least four days for equilibration. The reactions were monitored daily by LC-MS as described above for the single component library. In the case of templation studies, a second library was equilibrated in the presence of 25 mM of a five-residue polyproline peptide.

**vii. Photoisomerization.** Libraries were irradiated at 365 nm for five minutes in the dark with a Spectroline long wave UV pencil lamp (1,000 $\mu W/cm^2$ of 365 nm radiation at 1”). Once irradiated, solutions were continually stored in the dark at room temperature.

LC-MS analysis was continued on a daily basis and light exposure was minimized during analysis.

[1] (a) Miller, B. L.; Klekota, B. *Tetrahedron* **1999**, *55*, 11687-11697. (b) ten Cate, A. T.; Dankers, P. Y. W.; Sijbesma, R. P.; Meijer, E. W. *J. Org. Chem.* **2005**, *70*, 5799-5803.

[2] Leclaire, J.; Vial, L.; Otto, S.; Sanders, J. K. M. *Chem. Commun.* **2005**, 1959-1961.

[3] Rodriguez-Docampo, Z.; Otto, S. *Chem. Commun.* **2008**, 5301-5305.

[4] Goral, V.; Nelen, M. I.; Eliseev, A. V.; Lehn, J-M. *Proc. Nat. Acad. Sci.* **2001**, *98*, 1347-1352.

[5] Schultz, D.; Nitschke, J. R. *Angew. Chem. Int. Ed.* **2006**, *45*, 2453-2456.

[6] Nitschke, J. R.; Lehn, J.-M. *Proc. Natl. Acad. Sci.* **2003**, *100*, 11970-11974.

[7] Sarma, R. J.; Otto, S.; Nitschke, J. R. *Chem. Eur. J.* **2007**, *13*, 9542-9546.

[8] (a) Christinat, N.; Scopelliti, R.; Severin, K. *J. Org. Chem.* **2007**, *72*, 2192-2200. (b) Christinat, N.; Scopelliti, R.; Severin, K. *Angew. Chem. Int. Ed.* **2008**, *47*, 1848-1852.

[9] Eliseev, A. V.; Nelen, M. I. *J. Am. Chem. Soc.* **1997**, *119*, 1147-1148.

[10] Renner, C.; Moroder, L. *ChemBioChem* **2006**, *7*, 868-878.

[11] Kumar, G. S.; Neckers, D. C. *Chem. Rev.* **1989**, *89*, 1915-1925.

[12] For example, see: (a) Renner, C.; Kusebauch, U.; Loweneck, M.; Milbradt, A. G.; Moroder, L. *J. Peptide Res.* **2005**, *65*, 4-14. (b) Pieroni, O.; Fissi, A.; Angelini, N.; Lenci, F. *Acc. Chem. Res.* **2001**, *34*, 9-17. (c) Jurt, S.; Aemissegger, A.; Guntert, P.; Zerbe, O.; Hilvert, D. *Angew. Chem. Int. Ed.* **2006**, *45*, 6297-6300.

[13] (a) For a review, see: Natansohn, A.; Rochon, P. *Chem. Rev.* **2002**, *102*, 4139-4175. (b) Iftime, G.; Labarthet, F. L.; Natansohn, A.; Rochon, P. *J. Am. Chem. Soc.* **2000**, *122*, 12646-12650. (c) Tie, C.; Gallucci, J. C.; Parquette, J. R. *J. Am. Chem. Soc.* **2006**, *128*, 1162-1171.

[14] Harvey, J. H.; Trauner, D. *ChemBioChem* **2008**, *9*, 191-193.

[15] Shinkai, S.; Manabe, O. *Top. Curr. Chem.* **1984**, *121*, 67-104.

[16] Ueno, A.; Tomita, Y.; Osa, T. *Tetrahedron Lett.* **1983**, *24*, 5245-5248.

[17] Abell, A. D.; Jones, M. A.; Neffe, A. T.; Aitken, S. G.; Cain, T. P.; Payne, R. J.; McNabb, S. B.; Coxon, J. M.; Stuart, B. G.; Pearson, D.; Lee, H. Y-Y.; Morton, J. D. *J. Med. Chem.* **2007**, *50*, 2916-2920.

[18] For example, see: (a) Matsumoto, M.; Nicholas, K. M. *J. Org. Chem.* **2007**, *72*, 9308-9313. (b) Liu, J.; West, K. R.; Bondy, C. R.; Sanders, J. K. M. *Org. Biomol. Chem.* **2007**, *5*, 778-786. (c) Simpson, M. G.; Watson, S. P.; Feeder, N.; Davies, J. E.; Sanders, J. K. M. *Org Lett.* **2000**, *2*, 1435-1438. (d) Roberts, S. L.; Furlan, R. L. E.; Otto, S.; Sanders, J. K. M. *Org. Biomol. Chem.* **2003**, *1*, 1625-1633.

[19] Cousins, G. R. L.; Poulsen, S-A.; Sanders, J. K. M. *Chem. Commun.* **1999**, 1575-1576.

[20] For example, see: Bornaghi, L. F.; Wilkinson, B. L.; Kiefel, M. J.; Poulsen, S. A. *Tettrahedron Lett.* **2004**, *45*, 9281-9284.

[21] Nguyen, R.; Huc, I. *Chem. Commun*. **2003**, 942-943.

[22] (a) Aemissegger, A.; Krautler, V.; van Gunsteren, W. F.; Hilvert, D. *J. Am. Chem. Soc.* **2005**, *127*, 2929-2936. (b) A similar synthesis of the corresponding free acid of **4** was reported by Renner and coworkers: Dong, S-L.; Loweneck, M.; Schrader, T. E.; Schreier, W. J.; Zinth, W.; Moroder, L.; Renner, C. *Chem. Eur. J.* **2006**, *12*, 1114-1120.

[23] Ulysse, L.; Chmielewski, J. *Bioorg. Med. Chem. Lett.* **1994**, *4*, 2145-2146.

[24] Griffiths, J. *Chem. Soc. Rev.* **1972**, *1*, 481-493.

[25] Furlan, R. L. E.; Ng, Y.-F.; Cousins, G. R. L.; Redman, J. E.; Sanders, J. K. M. *Tetrahedron* **2002**, *58*, 771-778.

[26] Ball, L. J.; Kuhne, R.; Schneider-Mergener, J.; Oschkinat, H. *Angew. Chem. Int. Ed.* **2005**, *44*, 2852-2869.

[27] Crespo, L.; Sanclimens, G.; Montaner, B.; Perez-Tomas, R.; Royo, M.; Pons, M.; Albericio, Giralt, E. *J. Am. Chem. Soc*. **2002**, *124*, 8876-8883.

[28] (a) Corbett, P. T.; Sanders, J. K. M.; Otto, S. *Angew. Chem. Int. Ed*. **2007**, *46*, 8858-8861. (b) Severin, K. *Chem. Eur. J*. **2004**, *10*, 2565-2580.

[29] (a) Bulos, F.; Roberts, S. L.; Furlan, R. L. E.; Sanders, J. K. M. *Chem. Commun*. **2007**, 3092-3093. (b) Ludlow, R. F.; Liu, J.; Li, H.; Roberts, S. L.; Sanders, J. K. M.; Otto, S. *Angew. Chem. Int. Ed.* **2007**, *46*, 5762-5764. (c) Voshell, S. M.; Lee, S. J.; Gagne, M. R. *J. Am. Chem. Soc*. **2006**, *128*, 12422-12423.

[30] Gorostiza, P.; Volgraf, M.; Numano, R.; Szobota, S.; Trauner, D.; Isacoff, E. Y. *Proc. Nat. Acad. Sci*. **2007**, *104*, 10865-10870.

[31] Boulegue, C.; Loweneck, M.; Renner, C.; Moroder, L. *ChemBioChem* **2007**, *8*, 591-594.

# CHAPTER IV

# DYANMIC CYCLIC THIODEPSIPEPTIDE LIBRARIES FROM THIOL-THIOESTER EXCHANGE

(Reproduced, in part, with permission from Ghosh, S.; Ingerman, L. A.; Frye, A. G.; Lee, S. J.; Gagné, M. R.; Waters, M. L. *Org. Lett*. **2010**, *12*, 1860-1863.)

## A. Background and Significance

Since so many natural and synthetic cyclic peptides exhibit biological activity, cyclic peptides have gained much attention and are often viewed as ideal structures in the field of drug design.[1,2,3] For example, cyclosporin is a cyclic nonribosomal peptide that is widely used as an immunosuppressant, revolutionizing organ transplants.[1] In addition, both natural and synthetic cyclic peptides have recently been identified as HDAC inhibitors, which hold promise in the treatment of cancer.[4] As drug scaffolds, cyclic peptides are advantageous because they mimic native protein structures, exhibit enhanced metabolic stability, and are structurally preorganized, which reduces the entropic cost of binding.

Despite their therapeutic potential, the options for synthesizing structurally diverse libraries of cyclic peptides in a high throughput format are limiting.[3] For example, phage display has been used to generate and screen large libraries of disulfide-linked cyclic peptides as potential inhibitors.[5] However disulfides are not stable within the cellular

environment and there is no straightforward synthetic replacement for a disulfide linkage, limiting their practicality. Split-and-pool synthesis is also a popular method of generating libraries of cyclic peptides, however it requires orthogonal protecting group strategies for on-bead cyclizations and it can result in mixtures of both cyclic and linear peptides.[3]

Herein, we describe a strategy to use thioester exchange for generating solution-phase cyclic thiodepsipeptide libraries via dynamic combinatorial chemistry.[6] DCC allows for the *in situ* generation of complex mixtures of peptide macrocycles from smaller building blocks. The generation of cyclic peptide libraries via DCC is not unprecedented. A proline-phenylalanine dipeptide (Figure 4.1a) has been used to generate a series of peptidic macrocycles of varying sizes via hydrazone exchange, one of which has been demonstrated as an effective receptor for the neurotransmitter, acetylcholine.[7] Two different peptide macrocycles of this hydrazone DCL have also been shown to effectively recognize the diastereomeric templates quinine and quinidine.[8] Furthermore, a greater degree of diversity has been introduced into such peptide-based libraries through the investigation of monomers with the general building block structure shown in Figure 4.1b, where the residues in positions $AA_1$ and $AA_2$ are varied.[9]

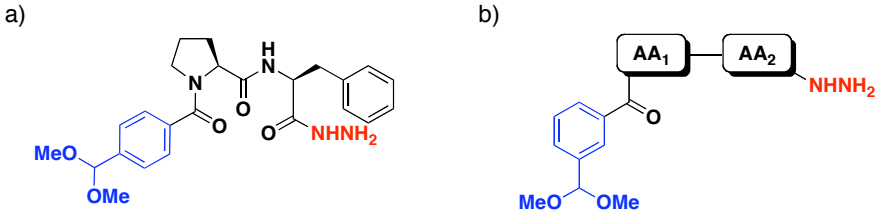



**Figure 4.1.** a) Proline-phenylalanine dipeptide monomer and b) the general building block structure used to investigate the relationship between peptide structure and library diversity through variation of positions $AA_1$ and $AA_2$.

While such reports of peptidic DCL libraries remain interesting, their relevance in the investigation of effective peptide receptors for biological targets of interest is debatable, as the peptide macrocycles are generated in a combination of organic solvents under very acidic conditions and contain an unnatural hydrazone linkage. A more desirable scenario for the investigation of potential drug scaffolds would be the generation of diverse libraries of cyclic peptides under close to physiological conditions. Recent reports suggest that thiol-thioester exchange is a promising reversible reaction for DCC since it is rapid in aqueous solution at neutral pH,[10,11] and it provides a native-like thioester linkage (Figure 4.2). The thioester moiety can also be subsequently replaced by more robust amide or ester functionalities, enhancing the potential utility of such macrocycles. However, commonly reports of thioester exchange of peptide based molecules make use of thioesters that are unsubstituted at the α-position, which significantly limits the structural diversity of the peptides that can be generated. As a result, we have investigated the reactivity of peptidic thiol-thioester monomers to determine the scope and limitations for their application to DCC.

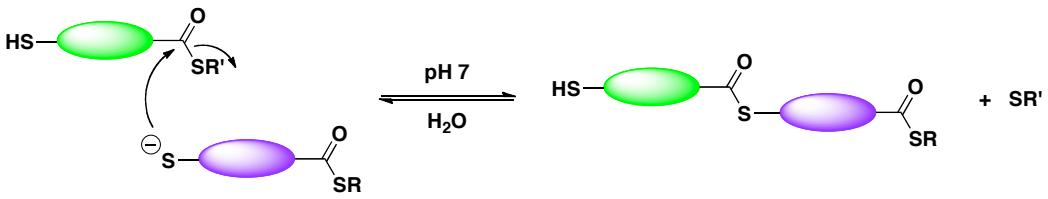
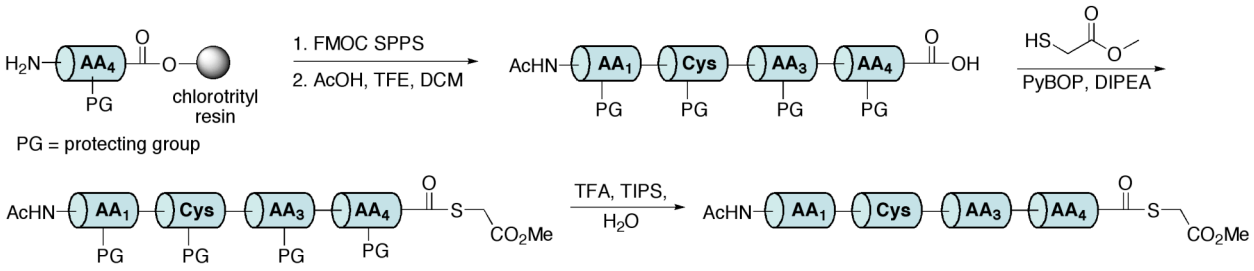



**Figure 4.2.** Reversible thioester exchange.

## B. Results and Discussion

**i. Design and synthesis.** Each peptide monomer used to investigate thiol-thioester exchange in the context of DCC libraries was designed as a four-residue peptide with the appropriate thiol and thioester functionality required to facilitate exchange, and the

general structure shown in Scheme 4.1. The peptides were prepared with a C-terminal thioester and a cysteine residue at position $AA_2$ to introduce the thiol functionality. The peptide monomers were synthesized using a modification of the method reported by Gellman and coworkers,[11] which provided all monomers in approximately 70 – 80% yield with no observable epimerization. This on-bead synthesis of the peptide monomers is also advantageous in that it allows for the facile generation of an array of monomers containing diverse functionality at various positions.




**Scheme 4.1.** Synthesis of thioester peptide monomers.

To introduce structural diversity into the DCLs, various amino acids were included at the $AA_4$ position to both investigate the effect on macrocyclization of having different side chains at the C-terminus, and to incorporate a variety of functional groups as potential recognition elements. Amino acids incorporated at the $AA_4$ position included positively charged amino acids (Lys, Arg), a negatively charged amino acid (Glu), a hydrogen bonding and neutral amino acid (Gln), and hydrophobic and sterically bulky amino acids (Phe, Val). Residues $AA_2$-$AA_3$-$AA_4$ were designed to form the peptide macrocycle, with $AA_1$ remaining exo-cyclic. Charged amino acids (Lys, Arg, Glu) were incorporated at position $AA_1$ to enhance water solubility of the macrocycles, as well as prevent aggregation. Similar to the previously developed hydrazone peptide monomers, a hydroxyproline turn residue was initially included at the $AA_3$ position to enforce a

degree of curvature and favor the formation of macrocycles, while the remainder of the structure provides flexibility.[12] Hydroxyproline was initially included to enhance water solubility, however proline was later investigated at the $AA_3$ position as well.

**ii. Preliminary DCLs: optimization and initial observations.** Multi-building block libraries were initially investigated to gauge the level of structural diversity that could be achieved in these DCLs and to probe the reactivity of these monomers. Early studies were focused on optimizing the reaction conditions for these libraries, such as the pH and buffer concentration. One principal advantage of thioester exchange is that exchange is generally complete within a few hours, in comparison to disulfide exchange which often requires weeks before equilibration is reached. However, if thioester equilibration is not reached within this time frame, the hydrolysis of both linear and cyclic thioesters becomes a competitive, irreversible process (Figure 4.3b). Therefore the reaction conditions must be optimized to ensure complete equilibration prior to the accumulation of hydrolyzed peptides.

Dynamic combinatorial libraries were prepared by mixing equimolar amounts of monomers **3**, **6**, and **7** with the general sequence Ac-Glu-Cys-Hyp-X-SR where X = Phe, Glu, and Val respectively (Table 4.1). Systematic studies were performed to investigate the monomer reactivity in a range of buffer concentrations including 10, 50, and 200 mM, while the pH of the buffer was also varied from 7 to 9. These studies were also performed in the presence of tris(2-carboxyethyl)phosphine (TCEP) to prevent competitive disulfide formation. In each case, the monomers underwent thiol-thioester exchange to form a mixture of peptide macrocycles (Figure 4.3a). The library

compositions were analyzed by LC-MS at various time points and all of the library members were easily identified.

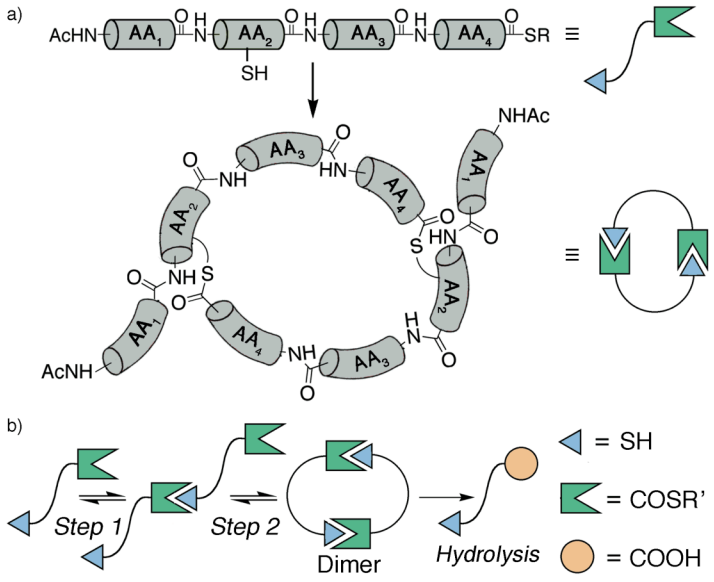



**Figure 4.3.** a) General design of monomers and the major cyclic product of the DCLs. b) Sequence of events toward dimeric cyclic thiodepsipeptides.

The monomer reactivity was found to be dependant on both the concentration and the pH of the buffer. Specifically, as both the buffer concentration and the pH increased, the rate of monomer disappearance increased as well. Very slow reactivity was observed with a buffer concentration of 10 mM, and after over a week significant amounts of unreacted starting material were still present in solution. While increasing the pH above neutral resulted in faster monomer reactivity, the rate of hydrolysis was seen to increase simultaneously, preventing library equilibration prior to the formation of significant concentrations of hydrolyzed products. It was therefore concluded that optimal thioester reactivity was achieved at a high buffer concentration (100 mM or greater), yet at a close to neutral pH. While the presence of a high concentration of TCEP successfully prevented the formation of disulfides, it also served to decrease the pH of the solutions

due to its multiple carboxylic acids, in turn resulting in a somewhat decreased rate of reactivity. As a result, TCEP was not included in later investigations of monomer reactivity, and degassing the buffer solutions was found to be equally effective in preventing disulfide formation.

While these early investigations did not thoroughly probe the potential structural diversity that could be achieved in these DCLs due to the presence of either unreacted starting material or hydrolyzed products, some interesting observations were made which may have otherwise gone unnoticed. For example, the analysis of a DCL containing monomers **3**, **6**, and **7** (Table 4.1) in 50 mM sodium phosphate buffer at pH 7.0 in the presence of an excess of TCEP revealed significantly different rates of disappearance of the three starting materials 1.5 hours after the library was generated (Figure 4.4). Monomers **3**, **6** and **7** were found to exist in a ratio of 1:1.7:2.5 at this time point, respectively, indicating the fastest reactivity for the N-terminal Phe monomer and the slowest reactivity for the C-terminal Val monomer. This reveals that the side chain of the residue at $AA_4$ appears to influence the rate at which the monomers undergo thioester exchange.

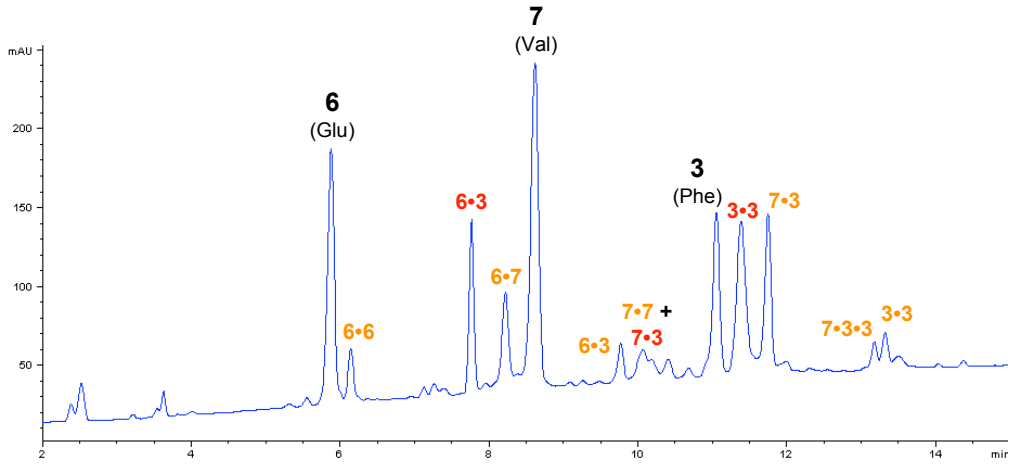



**Figure 4.4.** Analytical HPLC trace at 214 nm of a DCL consisting of monomers **3**, **6**, and **7** (2.5 mM each, in 50 mM sodium phosphate buffer with TCEP, pH 7.0) after 1.5 hours of equilibration, with the major species labeled. The three starting materials are labeled in black, the cyclic macrocycles are labeled in red, and the linear species are labeled in orange.

In addition, a mixture of both linear oligomer species and cyclic peptides were observed with this slow rate of reactivity. A simple two-step mechanism is proposed, where first the two monomers react to form a linear oligodimer, followed by a ring closing intramolecular transthioesterification reaction. Alternatively, if the oligodimer reacts with a third monomer prior to ring closing, a significant mount of larger macrocycles should be observed. However the library composition is dominated by dimeric species, indicating that the intermolecular step is likely rate limiting. While it appears that there has not been adequate time for cyclization of all of the oligodimers present in solution, monomer **3**, which appears to be disappearing the fastest, exists as one of the two components in each cyclic dimer formed, indicating that **3** indeed does appear to have a higher rate of macrocycle formation.

To probe this observation further, monomer **1** (Table 4.1) with an C-terminal glycine residue was investigated under identical conditions for a direct comparison, despite the

known slower reactivity under these conditions. If the side chain of $AA_4$ significantly influences the monomer reactivity, it would be expected that **1** would react just as fast if not faster than any other residue at this position. The analysis of this single building block library at various time points revealed that **1** in fact does react at a much faster rate than the monomers investigated previously. After a day **1** was almost completely consumed (Figure 4.5), whereas significant amounts of **3**, **6**, and **7** still remained unreacted. In addition, the library composition was limited to almost exclusively the dimeric 20-atom macrocyclic hexapeptide in this case, with only a trace amount of trimer. The lack of formation of an oligodimer in this case supports the observation that this monomer has a faster rate of macrocycle formation than **3**, and that the intermolecular step is rate limiting.

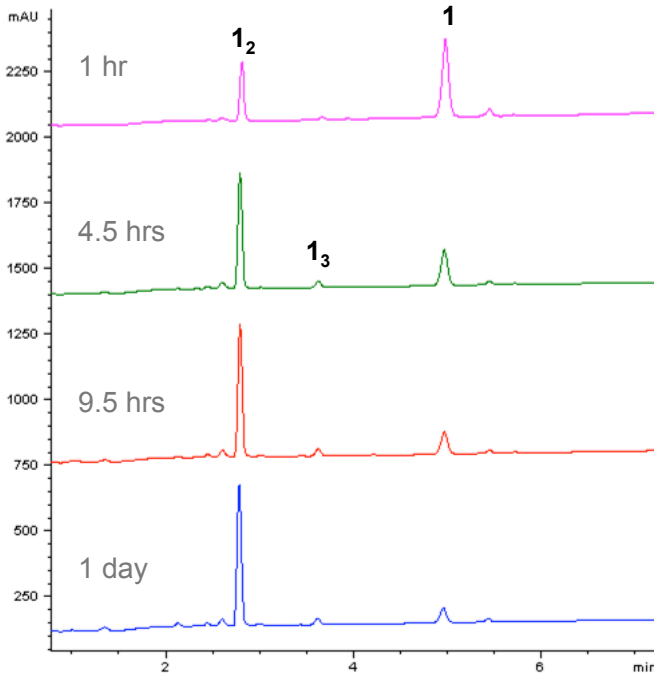



**Figure 4.5.** Analytical HPLC trace at 214 nm of a DCL consisting of monomer **1** only (2.5 mM **1** in 50 mM sodium phosphate buffer with TCEP, pH 7.0) after 1 hour (magenta), 4.5 hours (green), 9.5 hours (red), and 1 day (blue).

**iii. Systematic investigation of monomer reactivity.** To perform a thorough investigation of the influence of the C-terminal residue in the designed system on the rate of thiol-thioester exchange for the generation of peptide libraries, peptides **1** – **15** (Table 4.1) were investigated as single building block libraries under identical, optimal conditions. This rate study was largely conducted by Dr. Soumyadip Ghosh, but for completeness, will be included in this thesis. In addition, the effect of making subtle changes at the $AA_1$ and $AA_3$ positions on both the library composition and rate of reaction was considered. Since the intermolecular reaction is thought to be rate determining, the overall reaction rate was determined by monitoring the disappearance of monomer over time, as measured by a decrease in the monomer peak area in the HPLC trace (Figure 4.6). Since several monomers reacted too fast to accurately measure at pH 7, the reactivity was instead followed at pH 6.75. Importantly, at this pH, with the exception of monomers **7** and **15** containing Val at the C-terminus, equilibrium was always reached prior to the detection of hydrolyzed peptides.

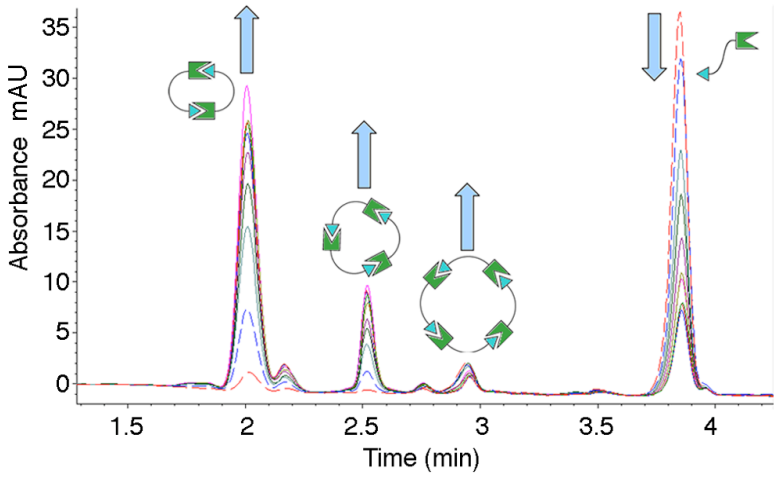



**Figure 4.6.** Overlay of analytical HPLC traces showing the consumption of monomer **5** (Ac-ECHypQ-COSR') and the appearance of cyclic macrocycles over time.

**Table 4.1.** Monomer sequences, half-lives of reactions, and libraries generated with **1** – **15** (100 mM $NH_4Ac$ Buffer pH 6.75, 25° C).

| compound | sequence | $t_{1/2}$ | product |
|---|---|---|---|
| **1** | Ac-ECHyp**G**-COSR' | ~ 12 min | dimer |
| **2** | Ac-ECP**H**-COSR' | ~ 12 min | dimer, trimer |
| **3** | Ac-ECHyp**F**-COSR' | ~ 1½ h | dimer, |
| **4** | Ac-EC**P**F-COSR' | ~ 1½ h | dimer |
| **5** | Ac-ECHyp**Q**-COSR' | ~ 1½ h | dimer, trimer, tetramer |
| **6** | Ac-ECHyp**E**-COSR' | ~ 5½ h | dimer |
| **7** | Ac-ECHyp**V**-COSR' | ND* | oligodimer |
| **8** | Ac-EC**W**E-COSR' | ~ 3½ h | dimer |
| **9** | Ac-EC(**D-P**)Q-COSR' | ~ 10 min | dimer, trimer |
| **10** | Ac-**K**CP**R**-COSR' | ~ 15 min | dimer |
| **11** | Ac-**R**CP**K**-COSR' | ~ 15 min | dimer |
| **12** | Ac-**K**CP**K**-COSR' | ~ 15 min | dimer |
| **13** | Ac-KCP**Q**-COSR' | ~ 20 min | dimer, trimer |
| **14** | Ac-KC**W**R-COSR' | ~ 15 min | dimer |
| **15** | Ac-KCP**V**-COSR' | ~ 10 h | dimer, trimer |

* Not determined. Monomers were hydrolyzed before complete reaction.

The C-terminal amino acid ($AA_4$) was varied systematically while holding the rest of the molecule constant (Table 4.1, compounds **1** – **3** and **5** – **7**). The observed reactivity trend, **1** ~ **2** < **3** ~ **5** < **6** < **7**, corresponds generally to differences in steric bulk at the C-terminal amino acid, with the reactivity decreasing with increasing steric bulk; however, in some cases, the differences in rate are more subtle (Figure 4.7). As expected, this trend is fairly consistent with the reactivity trends observed for native chemical ligation reactions of peptides and sugar-assisted ligation reactions of glycopeptides, which are initiated by the same first step involving attack of a thiolate anion.[13,14]

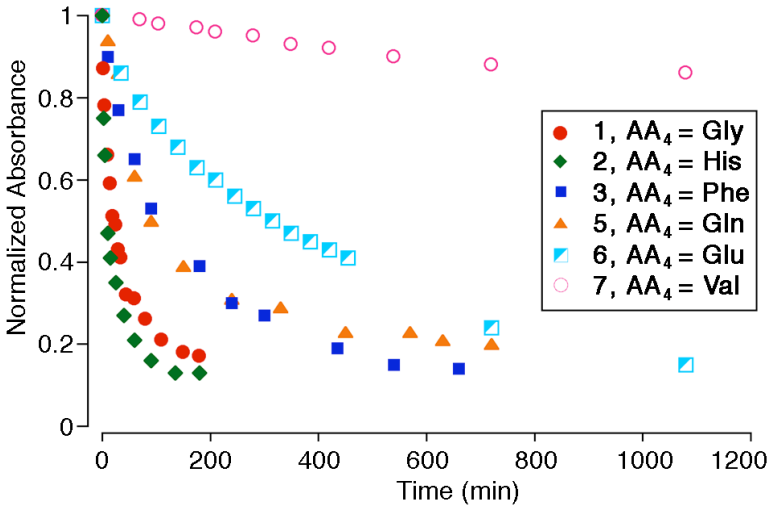



**Figure 4.7.** Plot of the disappearance of monomers **1** – **3** and **5** – **7** over time (100 mM $NH_4Ac$ buffer, pH 6.75, 25 °C).

It was observed that His reacts as rapidly as Gly, despite its larger size,[13] suggesting that the imidazole group of the C-terminal His stabilizes the transition state of the trans-thioesterification reaction. In contrast, β-branched amino acids such as Val react significantly more slowly (Figure 4.7), resulting in measurable hydrolysis prior to complete cyclization. Monomer **6**, which contains Glu at $AA_4$, reacts more slowly than its neutral Gln analog, while also forming multiple mass degenerate isomers with different retention times in the LC-MS analysis. On the basis of the MS analysis of individual peaks as well as literature precedent, it appears that the Glu γ-carboxyl cyclizes onto the thioester to form an anhydride, and ultimately **6-isom** (Figure 4.8).[15]

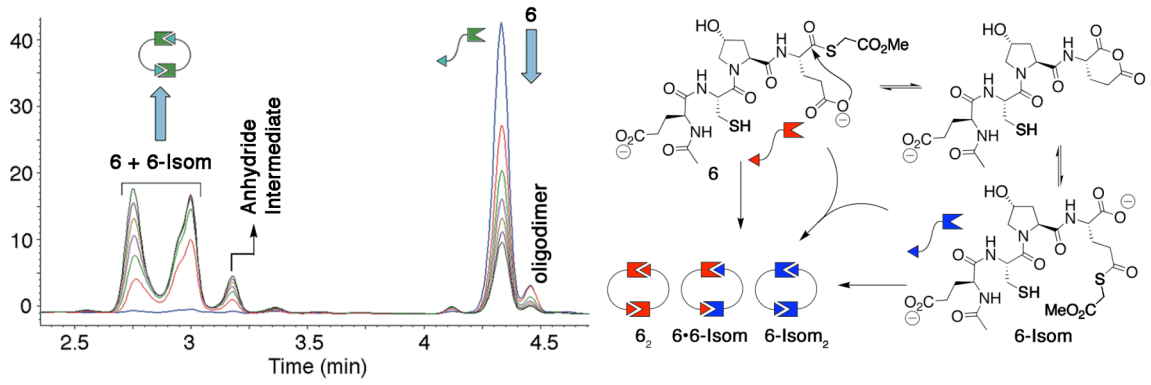



**Figure 4.8.** Reaction of the C-terminal Glu peptide **6** over time (left) and a proposed mechanism for the formation of side product **6-isom**, leading to multiple mass degenerate products.

To further explore the monomer reactivity, substitutions were made at positions $AA_1$ and $AA_3$ as well. Replacement of the Hyp in **3** with Pro resulted in identical HPLC traces and rate profiles for **3** and **4**, indicating that the OH group on hydroxyproline does not significantly affect the rate of exchange. The Hyp was also substituted with a more flexible Trp residue in **8** to determine if a turn residue is necessary to facilitate rapid cyclization. In fact, it was found that **8** reacts faster than its Hyp analog, and thus the rigidity of the proline turn may indeed impede reactivity. Stereochemistry was also determined to be a factor in the reaction rates, as replacement of L-Hyp (**5**) with D-Pro (**9**) resulted in considerably faster monomer reactivity. This result does not reflect conformational preferences for cyclization, but instead an influence of chirality on the accessibility of the thioester or thiol functional groups.

Lastly, the effect of positively charged amino acids (Lys and Arg) at positions $AA_1$ and $AA_4$ was evaluated. Interestingly, monomers **10** – **15**, which contain either Lys or Arg at $AA_1$, and in some cases at position $AA_4$ as well, reacted significantly faster than their negatively charged analogs. This rate enhancement suggests that the positively

charged Arg and Lys residues stabilize the build up of negative charge in the transition state through hydrogen bonding and/or electrostatic interactions. Furthermore, the Glu at $AA_1$ of the sterically hindered Val monomer **7** was replaced with Lys to give monomer **15**. Although hydrolysis was still observed before complete cyclization, **15** was found to react almost five times faster than **7** (Figure 4.9).

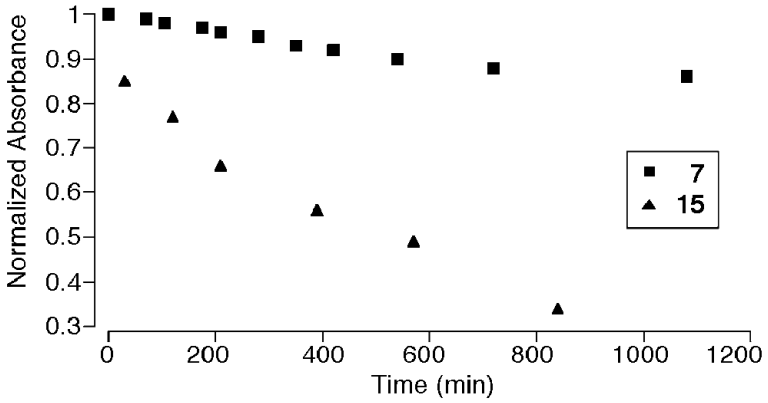



**Figure 4.9.** Comparison of the disappearance of monomers **7** and **15** over time, containing an N-terminal Glu and Lys, respectively (100 mM $NH_4Ac$ buffer, pH 6.75, 25 °C).

**iv. Generation of diverse libraries.** Having optimized the reaction conditions for rapid thiol-thioester cyclization with minimal hydrolysis, and subsequently evaluated the reactivity of a range of peptide monomers, we returned to the goal of generating structurally diverse libraries of cyclic peptides. By generating libraries containing multiple peptide monomers, the formation of macrocyclic thiodepsipeptides mixtures with homo and heterodimers as the major products were observed. To confirm thermodynamic control of the library composition, experiments were performed adding **1** to the preformed homodimer $\mathbf{3_2}$, and **3** to the preformed homodimer $\mathbf{1_2}$. Nearly superimposable HPLC traces were obtained after 18 h at pH 6.75, indicating that equilibrium can be reached from any direction.

It is clear that a high level of chemical diversity can be rapidly achieved by simply increasing the number of monomers combined in solution, as in this case by mixing **1**, **3**, and **5** – **7** (Figure 4.10). It is worth noting that monomer **7** was incorporated into the library despite its slow reactivity. In addition to cyclodimers, tripeptide macrocycles were also identified. The library speciation was readily deconvoluted using LC-MS, which showed extensive cross-reactivity of monomers and a general lack of self-sorting.[16] It also should be noted, however, that in the generation of even larger libraries of macrocyclic peptides, eventually the libraries become too complex and outgrow the capabilities of LC-MS analysis. This highlights the need for new methods of analysis which are able to accommodate libraries of great size, so that the potential of generating highly diverse and complex peptide libraries via DCC can be realized.

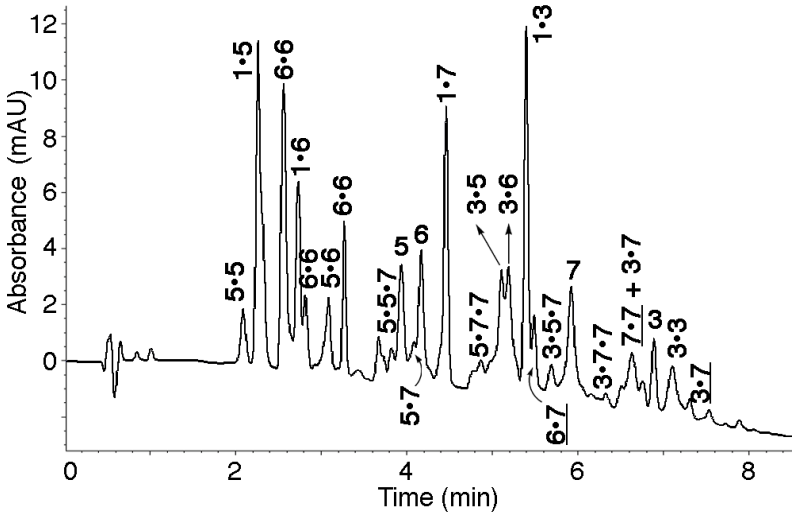



**Figure 4.10.** Analytical HPLC trace of a mixture of **1**, **3**, and **5**-**7** after equilibration for 1 hour (100 mM $NH_4Ac$ buffer, pH 6.75, 25 °C). Underlined species represent linear oligomers.

## C. Conclusion

In summary, an efficient method for rapidly generating a complex library of macrocyclic thiodepsipeptides at neutral pH for high-throughput screening has been demonstrated. Structure-function studies indicate that the thiol-thioester exchange reaction is tolerant of a variety of amino acids at the C-terminal position, with the exception of β-branched amino acids such as Val, which significantly retard the rate of homo-dimerization, and carboxylic acid side chains, which can isomerize through anhydride intermediates. Positively charged amino acids, Lys, Arg, and His, at the C-terminus were found to enhance the rate of thiol-thioester exchange, suggesting a stabilizing effect through hydrogen bonding and/or electrostatic interactions of the tetrahedral intermediate. Interestingly, positively charged amino acids were also found to increase reaction rates when incorporated at the N-terminus, suggesting that the transition state can be stabilized when flanked by a positively charged amino acid on either side of the peptide. In all cases, the dimer macrocycle is the major product, but the extent of formation of larger macrocycles is dependent on the peptide sequence. Upon mixing two or more monomers, we have shown that a complex library of cyclic thiodepsipeptides can be generated *in situ* under thermodynamic control.

This study demonstrates the feasibility of using DCC to generate a large number of cyclic thiodepsipeptides that can be efficiently screened against a particular target within hours. Since thioesters have transient stability *in vivo*, a receptor identified via a high-throughput screen can be easily re-synthesized as a more stable analog by replacing the thioester group with either an amide or an ester. Moreover, upon mixing monomers containing a Cys located at different positions, macrocycles of different sizes can in

theory be generated, such that the outcome is not limited to only cyclic hexapeptides. This idea of using cyclic peptide libraries generated via DCC for high throughput screens can be employed for a variety of applications, including the development of inhibitors, the disruption of protein-protein interactions, and the binding of nucleic acids.

## D. Experimental Section

**i. Peptide synthesis.** All peptides were synthesized by automated solid-phase synthesis in a Thuramed Tetras peptide synthesizer (version X94) using standard Fmoc chemistry on a 2-chlorotrityl resin (150 mg, apx. 0.09 mmol depending on resin loading). In each case, the resin was purchased with the C-terminal residue ($AA_4$) preloaded onto the resin. All amino acids with side chain functionality were protected during synthesis. Activation of amino acids was performed with HBTU and HOBT in the presence of DIEPA in DMF. Double coupling cycles (60 min each) were used for each amino acid coupling step. Fmoc deprotections were carried out in 2% DBU (1,8 diazabicyclo[5.4.0] undec-7-ene) and 2% piperidine in DMF. All peptide monomers were acetylated at the N-terminus with 5% acetic anhydride and 6% lutidine in DMF for 30 min. Cleavage of the peptide from the resin was performed in 8 mL of a 8:1:1 ratio of dry dichloromethane (DCM) : trifluoroethanol (TFE) : acetic acid (AcOH) for 1.5 h. The filtrate was concentrated by rotary evaporation to obtain the peptide product as a yellow oil. The product was dried under high vacuum and then dissolved in 3.6 mL of dry DMF. To that solution was added 32 μL of *N*,*N*-diisopropylethylamine (DIPEA), 18 μL methyl thioglycolate, and 86 mg PyBOP and the reaction was stirred under $N_2$ at room temperature for 1 h. DMF was removed under high vacuum to obtain a yellow oil. The

amino acid side chain protecting groups were removed in a solution of 18:1:1 trifluoroacetic acid (TFA) : triisopropylsilane (TIPS) : water for 4 hours. The TFA was evaporated and the cleavage products were precipitated with cold ether. The peptides were extracted into water, lyophilized to solid, and then purified by reverse-phase HPLC using a Vydac C-18 semipreparative column and a gradient of 0 to 80% B over 40 min, where solvent A was 95:5 water:acetonitrile, 0.1% TFA and solvent B was 95:5 acetonitrile:water, 0.1% TFA. After purification, the peptides were lyophilized to powder and characterized by +ESI-MS.

**ii. Thioester libraries.** The relevant building block solution(s) was first prepared at a higher concentration in water and diluted accordingly in buffer to give the desired final monomer concentration. Similarly, 50 – 250 mM buffer stock solutions were prepared and diluted accordingly to give the final desired buffer concentration. Any remaining volume was made up with water. The initial three component libraries (**3**, **6**, and **7**; 2.5 mM each) were prepared in sodium phosphate buffer with concentrations ranging from 10 – 200 mM at pH 7, 8, and 9, and 5 mM TCEP was included in each case. The single component libraries were prepared by making 1 mM monomer solutions in 100 mM $NH_4Ac$ buffer at pH 6.75. The buffer was degassed prior to use. The multicomponent libraries (> 3 monomers) were prepared by mixing equimolar amount of monomers (1 mM each) in 100 mM $NH_4Ac$ buffer at pH 6.75. The vials were capped and analyzed at various time points.

**iii. Analytical LC/MS.** LC-MS was carried out on an Agilent Rapid Resolution LC-MS system, equipped with an online degasser, binary pump, autosampler, heated column compartment, and diode array detector. Separations were performed on a Zorbax Eclipse

XDB-C18 column (4.6 × 50 mm, 1.8 micron). Separations were performed using either water-methanol gradients with 0.2% formic acid ($t$ = 0 min: 90% water, $t$ = 10 min: 0% water; flow rate = 1 mL/min) or water-acetonitrile gradients with 0.2% formic acid ($t$ = 0 min: 95% water, $t$ = 15 min: 62% water; flow rate = 1 mL/min), with the column temperature set to 45 ˚C to optimize the separation. Multiple wavelengths were monitored for analysis (214 nm, 230 nm, 254 nm, and 280 nm).

**iv. Monomer rate studies.** The reaction rate of each monomer was measured by analyzing the library composition at multiple time points. A 50 µl reaction aliquot was removed from the reaction in each case, and the reversible thiol-thioester exchange was quenched by adding it to a 50 µl solution of 9:1 water:TFA. The amount of monomer in each HPLC trace was determined by manual integration of the area under the monomer peak.

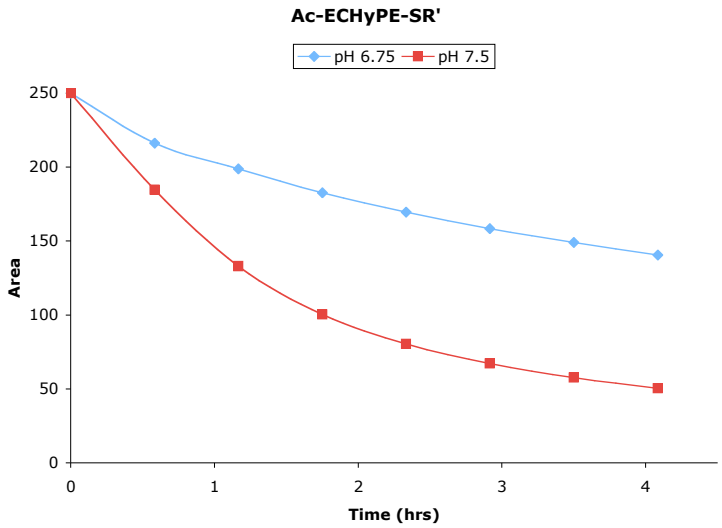



**Figure 4.11.** Plot of the decrease in concentration of **6** (1 mM) over time in100 mM $NH_4Ac$ buffer pH 6.75, RT (blue) and 100 mM K-Phosphate buffer pH 7.5, RT.

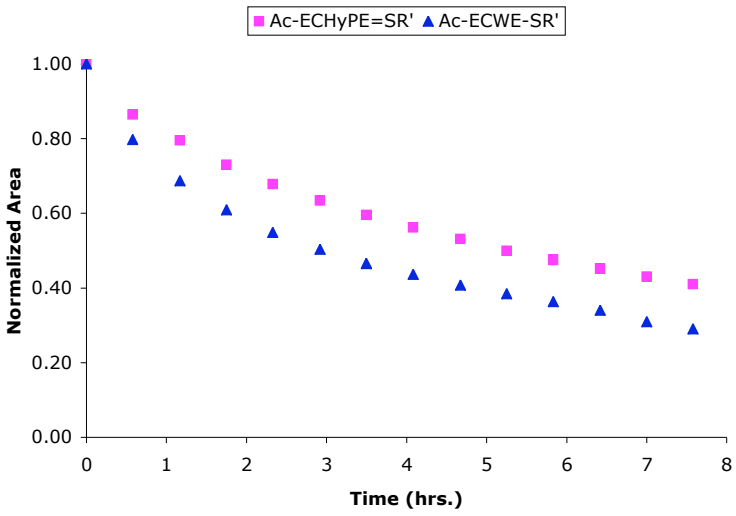



**Figure 4.12.** Plot of the decrease in concentration of monomers **6** (magenta) and **8** (blue) versus time. The monomer concentration is 1 mM (100 mM $NH_4Ac$ buffer pH 6.75, RT).

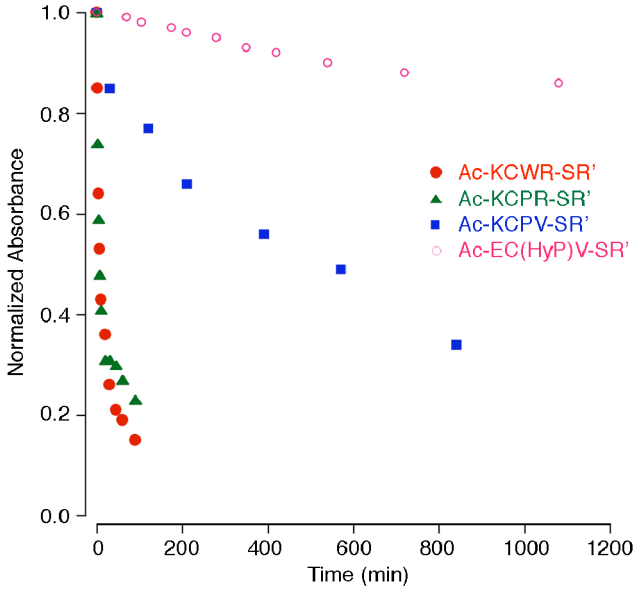



**Figure 4.13.** Plot of the decrease in concentration of monomers **7** (pink), **10** (green), **14** (red), and **15** (blue) over time (100 mM $NH_4Ac$ buffer pH 6.75, RT).

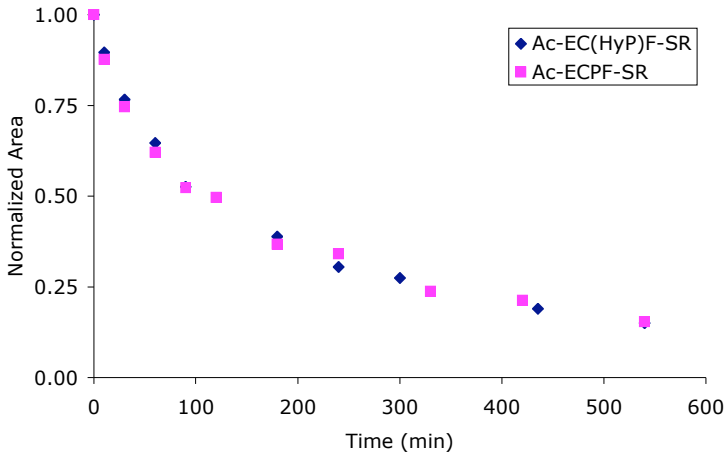



**Figure 4.14.** Plot of the decrease in concentration of monomers **3** (blue) and **4** (magenta) over time. The monomer concentration is 1 mM (100 mM $NH_4Ac$ buffer pH 6.75, RT).

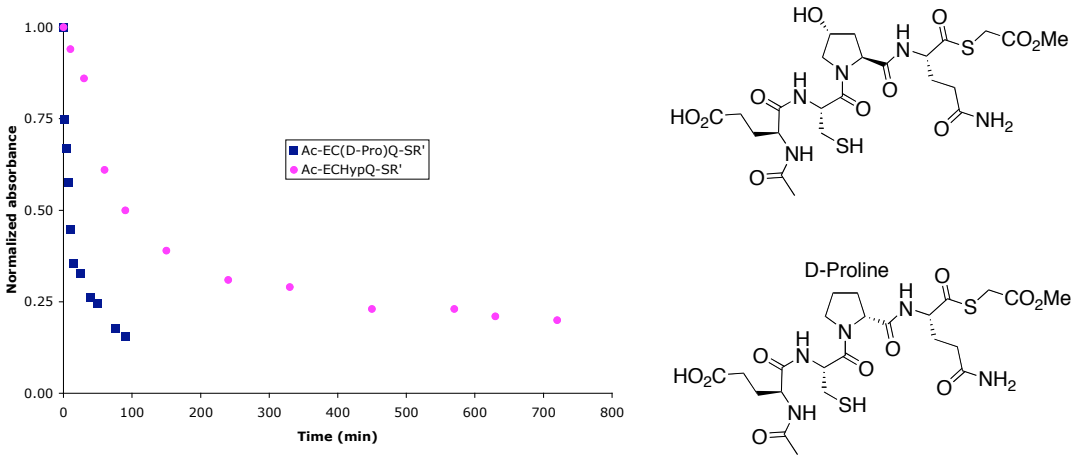



**Figure 4.15.** Plot of the decrease in concentration of monomers **5** (magenta) and **9** (blue) over time. The monomer concentration is 1 mM (100 mM $NH_4Ac$ buffer pH 6.75, RT).

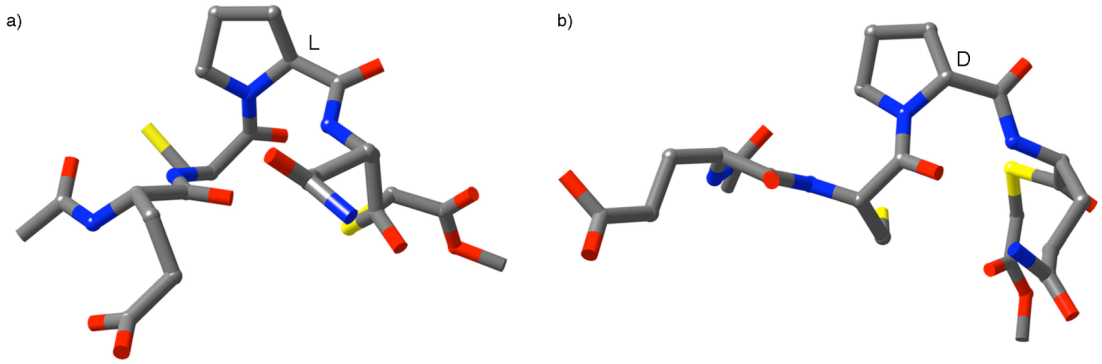



**Figure 4.16.** Molecular models of (a) **5** and (b) **9**. Hydrogen atoms are removed for clarity. Color coding: C, grey; N, blue; O, red; and S, yellow.

[1] Craik, D. J. *Science* **2006**, *311*, 1563-1564.

[2] Horswill, A. R.; Benkovic, S. J. *Cell Cycle* **2005**, *4*, 552-555.

[3] Horton, D. A.; Bourne, G. T.; Smythe, M. L. *J. Comput. Aided Mol. Des.* **2002**, *16*, 415-430.

[4] (a) Minucci, S.; Pelicci, P. G. *Nat. Rev. Cancer* **2006**, *6*, 38-51. (b) Meinke, P. T.; Liberator, P. *Curr. Med. Chem.* **2001**, *8*, 211-235. (c) Shute, R. E.; Kawai, M.; Rich, D. H. *Tetrahedron* **1988**, *44*, 685-95. (d) Montero, A.; Beierle, J. M.; Olsen, C. A.; Ghadiri, M. R. *J. Am. Chem. Soc.* **2009**, *131*, 3033–3041. (e) Olsen, C. A.; Ghadiri, M. R. *J. Med. Chem*, **2009**, *52*, 7836–7846.

[5] Sidhu, S. S.; Fairbrother, W. J.; Deshayes, K. *ChemBiochem*, **2003**, *4*, 14-25.

[6] Corbett, P. T.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J. L.; Sanders, J. K.; Otto, S. *Chem. Rev.* **2006**, *106*, 3652-3711.

[7] Lam, R. T. S.; Belenguer, A.; Roberts, S. L.; Naumann, C.; Jarrosson, T.; Otto, S.; Sanders, J. K. M. *Science* **2005**, *308*, 667-669.

[8] Bulos, F.; Roberts, S. L.; Furlan, R. L. E.; Sanders, J. K. M. *Chem. Commun.* **2007**, 3092-3093.

[9] Liu, J.; West, K. R,; Bondy, C. R.; Sanders, J. K. M. *Org. Biomol. Chem.* **2007**, *5*, 778-786.

[10] (a) Leclaire, J.; Vial, L.; Otto, S.; Sanders, J. K. M. *Chem. Commun.* **2005**, 1959-1961. (b) Larsson, R.; Pei, Z.; Ramström, O. *Angew. Chem. Int. Ed.* **2004**, *43*, 3716-3718. (c) Ura, Y.; Beierle, J. M.; Leman, L. J.; Orgel, L. E.; Ghadiri, M. R. *Science* **2009**, *325*, 73-77.

[11] Woll, M. G.; Gellman, S. H. *J. Am. Chem. Soc.* **2004**, *126*, 11172-11174.

[12] (a) Cousins, G. R. L.; Poulsen, S.-A; Sanders, J. K. M. *Chem. Commun.***, 1999**, 1575-1576. (b) Roberts, S. L.; Furlan, R. L. E.; Otto, S.; Sanders, J. K. M. *Org. Biomol. Chem.* **2003**, *1*, 1625-1633.

[13] Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. *Proc. Natl. Acad. Sci.* **1999**, *96*, 10068-10073.

[14] Brik, A.; Yang, Y. Y.; Ficht, S.; Wong, C. H. *J. Am. Chem. Soc.* **2006**, *128*, 5626-5627.

[15] Villain, M.; Gaertner, H.; Botti, P. *Eur. J. Org. Chem.* **2003**, 3267-3272.

[16] (a) Rowan, S. J.; Hamilton, D. G.; Brady, P. A.; Sanders, J. K. M. *J. Am. Chem. Soc.*, **1997**, *119*, 2578–2579. (b) Rowan, S. J.; Reynolds, D. J.; Sanders, J. K. M. *J. Org. Chem.* **1999**, *64*, 5804-5814. (c) Nitschke, J. R. *Acc. Chem. Res.*, **2007**, *40*, 103–112.

# CHAPTER V

# SELECTIVE RECOGNITION OF 7-METHYL GMP BY MACROCYCLIC PEPTIDE RECEPTORS IDENTIFIED BY DYNAMIC COMBINATORIAL CHEMISTRY

## A. Background and Significance

**i. β-Hairpin model systems for the recognition of GTP.** Short peptide sequences that adopt a specific secondary structure are often used as model systems to mimic a certain protein fold and study the factors that contribute to protein folding and protein recognition. Motivated by the prevalence of β-sheet motifs throughout protein structures and the desire to better understand the factors contributing to their stability, short, monomeric β-hairpin peptides have been studied as models for antiparallel β-sheets.[1] These β-sheet mimics contain two antiparallel β-strands connected by a turn sequence which nucleates the hairpin structure.[2] This two amino acid sequence imparts a twist to the β-turn, which in combination with the right-handed twist of the strands, results in a stable β-sheet.

Much of the work in this field has been focused on determining the factors that contribute to the stability of the hairpin fold, such as the β-turn sequence, sidechain-sidechain interactions, and individual amino acid β-sheet propensities.[3] These model systems serve largely as a framework for studying the non-covalent interactions that guide protein folding, ranging from hydrogen bonding to the formation of salt-bridges,[4] as well as aromatic[5] and cation-π interactions.[6]

It has also been found that β-hairpin peptides can serve as novel molecular receptors for biologically relevant aromatic compounds, such as nucleotides. Hairpins are quite valuable receptors for molecular recognition, as they mimic native protein binding motifs, have a large surface area, are well folded without covalent constraints, and allow for the incorporation of a wide range of functional groups. The binding sites of nucleotides, such as adenosine 5'-triphosphate (ATP), in proteins generally consist of aromatic, aliphatic, cationic, and hydrogen bonding residues.[7] The adenine moiety of ATP is typically found stacked with aromatic side chains, while the phosphates interact with the side chains of lysine or arginine.

These interactions were demonstrated in the β-hairpin peptide **WKWK** (Figure 5.1) that binds a molecule of ATP strongly in water via aromatic and electrostatic interactions on one face of the β-hairpin.[8] The diagonal Trp-Trp pair in the non-hydrogen bonding sites of the hairpin provides a binding cleft for aromatic intercalation (Figure 5.2). In addition, the lysine side chains on the same face of the hairpin provide electrostatic interactions with the ATP phosphate groups. ATP was found to bind the β-hairpin with a dissociation constant of 170 μM by fluorescence quenching, and a total binding energy of approximately -5 kcal/mol. By NMR titration, the dissociation constant was measured to be 1.4 mM as determined by the upfield shifting of the Trp protons upon binding. This 8-fold decrease in affinity is presumably due to the presence of 10 mM sodium chloride, which screens for favorable electrostatic interactions.

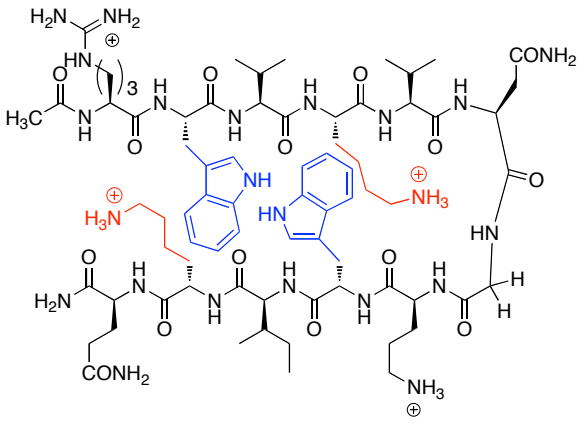


**Figure 5.1.** Structure of peptide **WKWK** with the sequence Ac-RWVKVNGOWIKQ-$NH_2$.

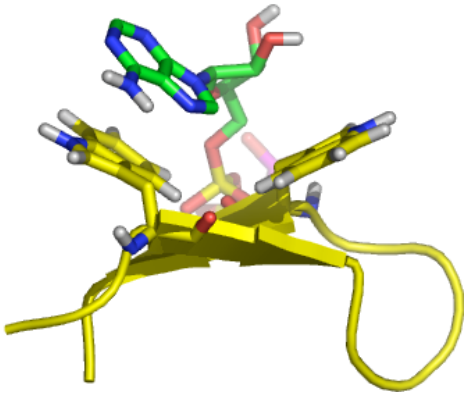


**Figure 5.2.** NMR structure of the adenine ring of ATP (green) bound to the β-hairpin WKWK (yellow) through intercalation of the two Trp residues which create an aromatic cleft for binding.

When nucleotide selectivity was evaluated by NMR, binding saturation was apparent at lower concentrations of GTP as compared to ATP, resulting in a $K_d$ of 0.45 mM and demonstrating a modestly higher affinity for this substrate. This selectivity could be related to the carbonyl in the guanine base, which has the potential to hydrogen bond with the flanking Lys residues (Figure 5.3). In both cases, upfield shifting of the ribose protons suggests that CH-π interactions with the ribose ring may also contribute to binding. These observations enhance the general understanding of protein-nucleic acid

interactions that occur with β-sheet proteins, while the observed selectivity for GTP shows promise for the development of protein receptors that are selective for guanosine-rich DNA sequences.

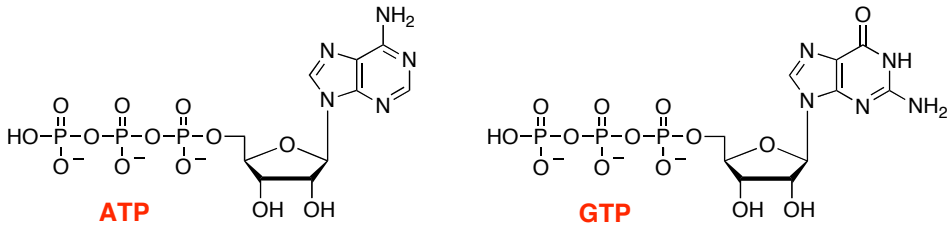



**Figure 5.3.** Structures of nucleotide substrates ATP and GTP.

**ii. Significance of 7-methyl GTP.** In addition to the natural bases, DNA is also known to contain methylated bases, and the biological implications of nucleic acid methylation is most commonly related to the control of gene expression.[9] In particular, eukaryotic mRNA is modified by the addition of the 7-methylguanosine 'cap' to the first transcribed nucleotide, and this modification is necessary for efficient gene expression and cell viability (Figure 5.4). The 7-methylguanosine cap is required for translation of the majority of mRNAs, and it has also been reported to stabilize mRNA against attack by exonucleases, as well as promote transcription, splicing, polyadenylation, and nuclear export of mRNA.[10] Formation of 7-methylguanosine occurs as the substrate mRNA is being transcribed, and it is catalyzed by enzymes which are recruited to RNA polymerase II. Furthermore, cellular proteins which regulate mRNA cap methylation are likely to function either by regulating the recruitment of the cap methyltransferase to RNA polymerase II, or by regulating the cap methyltransferase activity.[11]

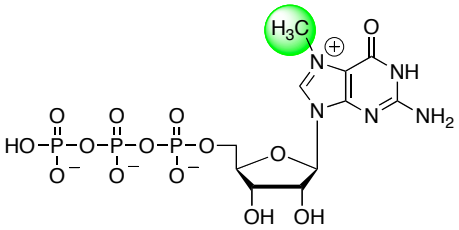



**Figure 5.4.** Structure of 7-methylguanosine triphosphate.

During the initiation of translation, the cap is specifically recognized by eukaryotic Initiation Factor 4E (eIF4E), which participates in the recruitment of ribosomes to mRNA.[12] This protein plays a crucial role in the translational control of gene expression, and it is known to be overexpressed in many types of tumor cells.[13] The critical role of translation initiation in cell growth and malignant transformation, coupled with the presence of abnormally high levels of eIF4E in several human cancers, suggests that the eIF4E-cap interaction could be a potential target for the development of anti-cancer drugs.[14] Consequently, gaining a greater understanding of this nucleotide-protein interaction would be quite valuable.

The specific recognition of the cap by eIF4E is thought to be largely mediated by the formation of a stacked configuration between the 7-methylguanine and two tryptophan indole rings, allowing for discrimination of the methylated cap over other unmethylated nucleotides. This association has been demonstrated in model systems between analogues of the mRNA cap and both indole derivatives and tryptophan-containing peptides.[15] It has also been proposed that a positive charge localization at the N7 position of the guanine ring is the primary reason for the enhanced stacking with the cap binding site of the protein, and leads to a closer analogy between this cation-π stacking interaction and the interaction common in protein structures between an aromatic ring of

one residue and a positively charged side chain of another.[16,17] This stacking between cationic 7-methylguanosine and the Trp side chains is necessary to effectively stabilize the base moiety, which in turn facilitates the formation of hydrogen bonds at the cap binding site.[18] Furthermore, aromatic stacking of the 7-methylguanine ring with two aromatic side chains seems to be a general mode of recognition of the mRNA 5' cap by various proteins.

**iii. Goal of this work.** While the recognition of nucleotides by β-hairpin model systems has been demonstrated, a system that exhibits an even greater affinity for such nucleotides would be desirable for the broader goal of DNA recognition. This could potentially be achieved by increasing the complexity of the hairpin receptor by various means, such as adding a third peptide strand to increase the number of contacts with the nucleotide, introducing unnatural amino acids with superior recognition properties, or by converting the hairpin binding cleft into a tighter binding pocket. In this work, we investigate dynamic combinatorial chemistry as a method for identifying novel peptide receptors for nucleotides. By introducing a covalent tether between the two antiparallel strands of the known receptor, WKWK, a tighter nucleotide binding pocket may be formed while increasing the potential number of favorable interactions. Furthermore, DCC allows for the identification and optimization of such peptide receptors in a facile and high-throughput manner. This system also facilitates an interesting investigation of the interplay between nucleotide binding and peptide folding, as a covalent modification of the hairpin to reinforce binding has the potential to disturb the hairpin fold.

In addition, due to the need for a greater understanding of the recognition of the 7-methylguanoisine mRNA cap by protein receptors, 7-methyl GMP will also be

investigated in this system, particularly in comparison to its unmethylated GMP counterpart. The presence of two Trp residues in the native WKWK sequence, as well as the potential to introduce further aromatic recognition elements via DCC, provides an opportunity to closely mimic native 7-methylguanosine protein receptors in this model system. In addition to highlighting the identification of novel nucleotide peptide receptors, these studies also allowed for preliminary investigations of the use of hairpin peptides as monomers in dynamic combinatorial libraries.

## B. Results and Discussion

**i. General library design.** In order to facilitate the formation of a tighter binding pocket for nucleotide recognition via DCC, two types of monomers were designed for reversible exchange: hairpin peptide monomers and small molecule monomers intended to bridge the two antiparallel strands of the peptide. These building blocks can be reversibly linked under thermodynamic control to produce an equilibrium mixture of potential nucleotide receptors. While the reaction of peptide monomers with small molecule monomers is desired for these purposes, the formation of hairpin dimers or larger peptide oligomers is also possible, which have the potential to serve as equally valuable nucleotide receptors.

Disulfide exchange has become one of the most widely used reactions in dynamic combinatorial libraries, and it is particularly suited for this biological application as it occurs in aqueous solution at close to neutral pH. Disulfide exchange is advantageous in that disulfides form readily from thiols in the presence of oxygen and small amounts of base, it takes place under mild conditions in the presence of a catalytic amount of thiol, it

can be quenched under acidic conditions, and disulfides are stable toward many different functional groups.[19] During oxidation the mixture contains both disulfides and thiols, allowing for equilibration through nucleophilic attack of thiolate anions on the disulfides, displacing a new thiolate anion in the process (Figure 2.8). The primary downside to disulfide exchange is that it occurs relatively slowly, sometimes requiring on the order of weeks for equilibration to be reached.

The incorporation of thiols into the hairpin structure was easily facilitated by the replacement of two residues of the native WKWK sequence with cysteine residues. Two variants of the dithiol hairpin were synthesized, peptides **1** and **2**, with two of the non-hydrogen bonding residues replaced with cysteines in each case (Figure 5.5). Non-hydrogen bonding residues were chosen so that exchange with the small aromatic monomers would occur on the same face of the hairpin as the Trp and Lys residues. While this results in the loss of one Trp and one Lys residue in each case which were shown previously to assist in nucleotide recognition, it is anticipated that new interactions will be generated to compensate for such losses. Despite their similarities, it is possible that the two hairpin monomers will function quite differently, as the introduction of an unnatural bridge located further from the turn in **2** is less likely to disrupt the hairpin structure than when introduced closer to the turn in **1**. Both peptides were synthesized by Fmoc solid phases peptide synthesis.

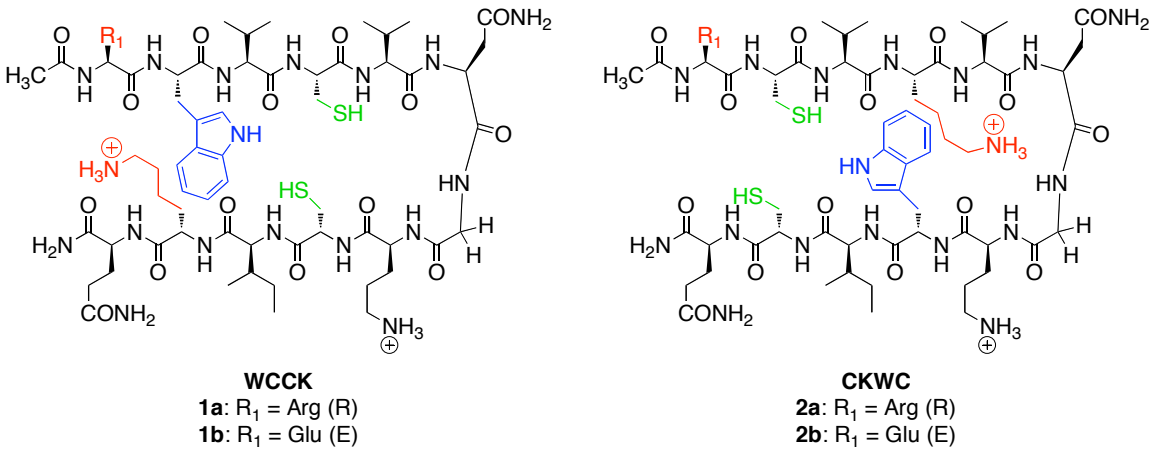



**Figure 5.5.** Hairpin dithiol monomers **1** and **2**, containing two Cys residues (green) in place of Trp and Lys residues of the native WKWK sequence. The remaining Trp residue is shown in blue, and the lysine in red.

Small molecule dithiols **B** and **C** were chosen as potential bridging groups for the designed dithiol hairpins, both of which have been discussed previously in chapter two (Figure 5.6). They include an aromatic surface to allow for aromatic and hydrophobic interactions with nucleotide bases, as well as potential cation-π interactions with 7-methylguanosine. The carboxylates are required to induce water solubility, although they also have the potential to participate in favorable electrostatic interactions or hydrogen bonds with the nucleotides. It should also be noted that the carboxylates have the potential to create repulsive interactions with the phosphate tails, if placed in close enough proximity. As the optimal length required to favorably bridge the two peptide strands in the designed positions is unknown, the distance between the thiols in **B** and **C** was varied, with monomer **C** containing two extra methylene groups in comparison to **B**, as well as an increased degree of flexibility. Both molecules also contain some degree of curvature, which should encourage the formation of cycles macrocycles with **1** or **2**.

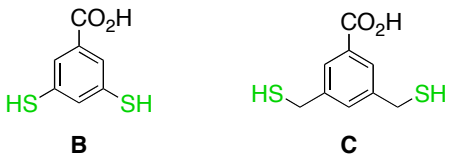



**Figure 5.6.** Structures of small aromatic dithiols, **B** and **C**.

Initial investigations of peptides **1a** and **2a** containing an N-terminal arginine in combination with **B** and **C** under the basic conditions required for disulfide exchange led to significant amounts of precipitation during equilibration. Insoluble macrocycles in a DCL act as a thermodynamic trap, shifting the equilibrium for reasons other than favorable binding. As both **1a** and **2a** in solution alone show no signs of insolubility over extended periods of time, this precipitation was clearly the result of interactions, either covalent or non-covalent, between the hairpin peptide and small aromatic monomers. Complete solubility could be regained with the addition of organic cosolvents, however it was anticipated that the use of co-solvents would prove to be limiting in the pursuit of future templation studies.

With an N-terminal arginine, each peptide has 3 basic residues. While an increase in the number of charged residues generally enhances peptide solubility, in combination with various equivalents of **B** and **C**, each of which contains a negative carboxylate, the charge of the library as a whole is close to neutral. By mutating the arginine, which is far removed from the intended binding site, to a glutamic acid (**1b** and **2b**), decreasing the overall peptide charge to one, these peptides no longer exhibited solubility problems in untemplated DCLs with monomers **B** and **C** at millimolar concentrations.

**ii. Hairpin monomer generation 2: $^{D}$Pro-Gly turn.** Initial investigations of untemplated DCLs containing either **1b** or **2b** as well as **B** and **C** resulted in a diverse

library of species as determined by HPLC. Interestingly, in the identification of the library members by LC-MS, it was found that various species were present in solution that were identified to be 1 mass unit greater than expected (Table 5.1), which is outside the error of the mass spectrometer. In each case, the presence of a peak at a different retention time with the expected mass indicated that this extra mass unit was indeed the result of a chemical change.

**Table 5.1.** Expected and observed masses for representative cyclic macrocycles in the DCL.

| Identity | Actual $[M+H]^+$ | Observed $[M+H]^+$ |
|---|---|---|
| **1b** or **2b** | 1433.7 | -- |
| **cyclic 1b** or **2b** (internal S-S) | 1431.7 | 1431.7 and 1432.7 |
| **(1b** or **2b) + B** | 1615.7 | 1615.7 and 1616.7 |
| **(1b** or **2b) + C** | 1643.7 | 1643.7 and 1644.7 |

It was proposed that deamidation of the asparagine turn residue was occurring, which results in an overall mass increase of 1 Da due to the conversion of asparagine to aspartic acid (Figure 5.7). This most commonly occurs with solvent accessible Asn-Gly sequences, such as the turn sequence of **1** and **2**, as exposure to alkaline pH results in an increased rate of deamidation due to the greater nucleophilicity of the backbone nitrogen adjacent to the Asn residue.[20] As these peptides are required to stir in alkaline solution for days and even weeks at a time to reach equilibrium, this side reaction is not unlikely.

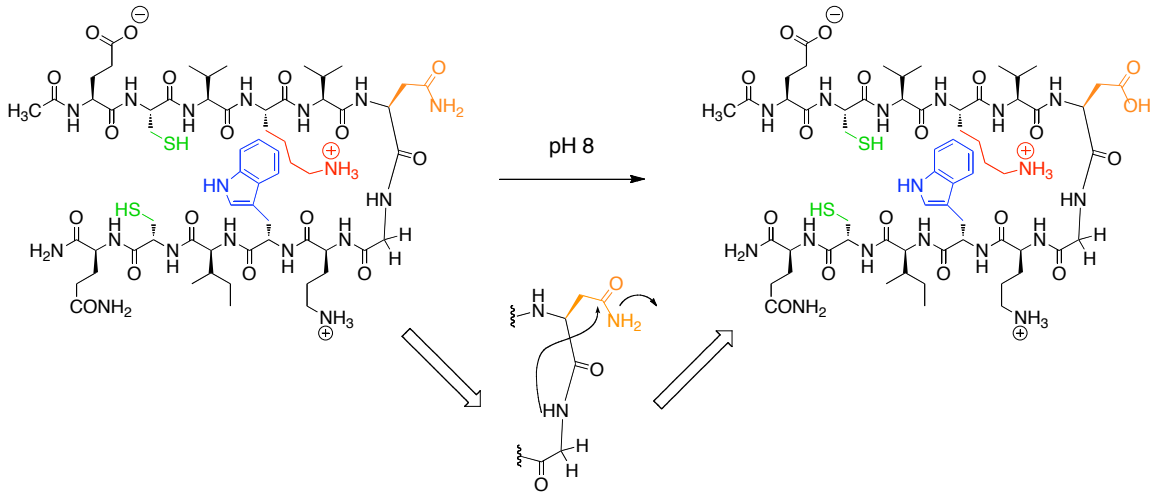
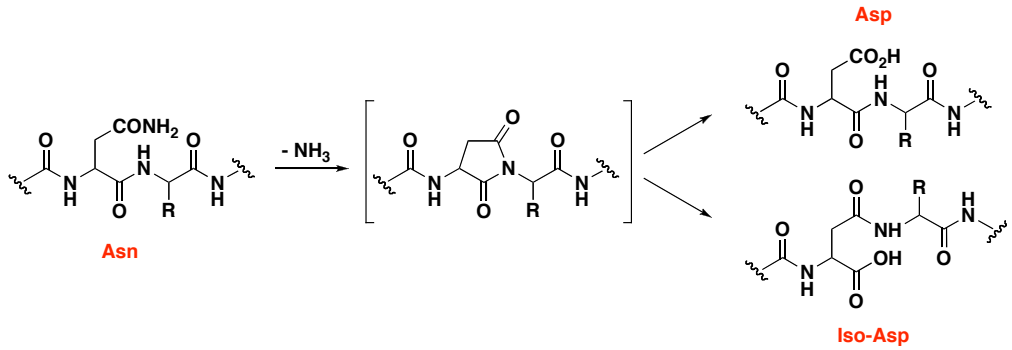


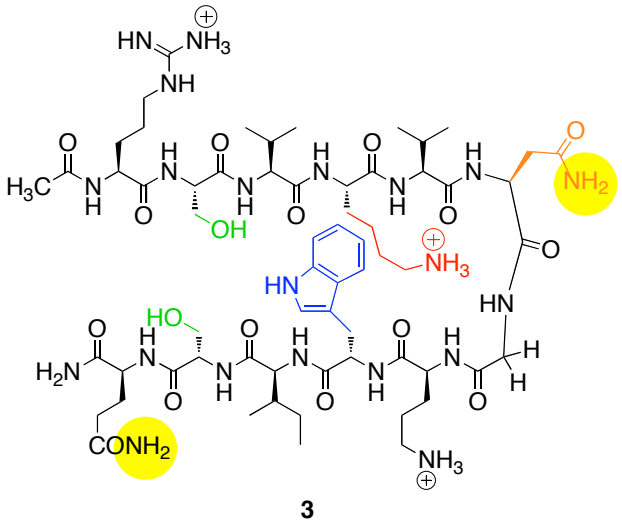

**Figure 5.7.** Deamidation of hairpin **2b** via attack of the carbonyl of the asparagine side chain (orange) by the backbone amide nitrogen.

While some peptide structures can tolerate such a slight modification, deamidation is viewed as an undesirable side reaction at the turn of a β-hairpin due to the fact that it disrupts the hairpin conformation. The conversion of asparagine to aspartic acid through a succinimide intermediate also allows for the conversion of asparagine to isoaspartic acid, which results in an even larger structural change of the hairpin peptide (Scheme 5.1).[21] IsoAsp is found to be the dominant of the two products, formed in a ratio of about 4:1 in basic solution.




**Scheme 5.1.** Deamidation of asparagine to aspartic acid or isoaspartic acid via a succinimide intermediate.

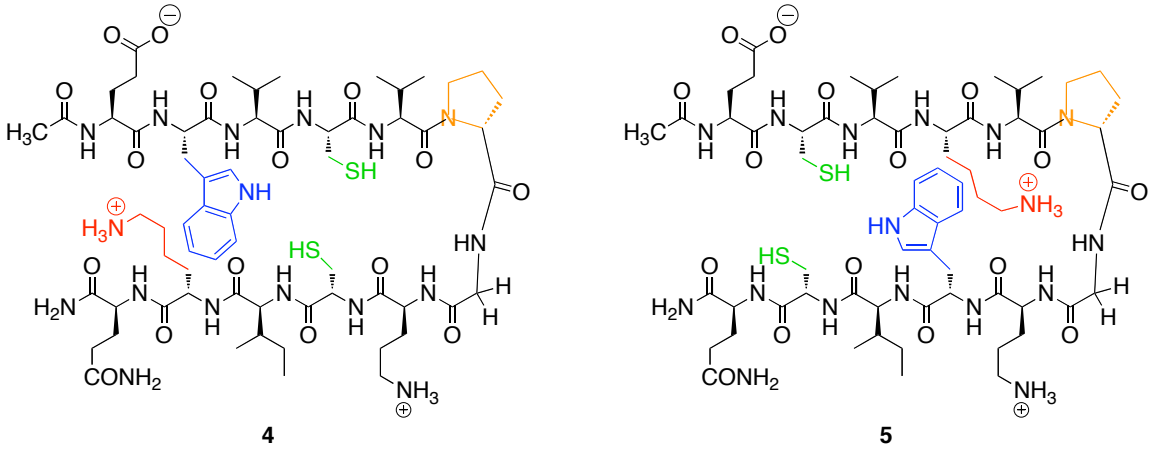
Deamidation of the hairpin turn was confirmed by NMR studies. Peptide **3** was synthesized (Figure 5.8), where the cysteine residues of **2b** were replaced with serine residues to avoid spectra complicated by the formation of multiple species via intermolecular and intramolecular disulfide bonds. The peptide was analyzed by NMR prior to and after being stirred in aqueous solution at pH 8-9 for three days. To assess the extent of peptide deamidation, the ratio of the asparagine side chain amide proton to the C-terminal glutamine side chain amide proton was evaluated. As the Gln side chain is unaffected by deamidation, the ratio should remain constant if this process is not occurring. However, the ratio of the Asn NH proton to the Gln NH proton decreased from 1.24:1 to 0.39:1 upon exposure to basic solution as determined by integration, suggesting the hairpin was indeed undergoing deamidation. The glycine splitting was also examined in each case, as the separation of the Gly $H_\alpha$ resonances have been shown to correlate with the extent of folding in β-hairpins.[6] While the glycine protons revealed a splitting pattern characteristic of a well folded system in the initial peptide NMR, following exposure to base, new proton signals were observed in the region of the glycine protons (3.5 – 4.0 ppm) indicating the presence of new splitting patterns and therefore a change in structure.



**3**

**Figure 5.8.** SKWS hairpin peptide **3** used to investigate asparagine deamidation by NMR. The ratio of the Gln and Asn side chain amide protons was investigated (highlighted in yellow).

Deamidation was easily prevented in future studies through the synthesis and use of peptides **4** and **5**, where the original Asn-Gly turn was replaced with a similar type I' $^{D}$Pro-Gly turn (Figure 5.9). The $^{D}$Pro-Gly turn has been shown to be equally, if not more effective in nucleating the formation of β-hairpins, and therefore should not affect the overall peptide stability.[22]

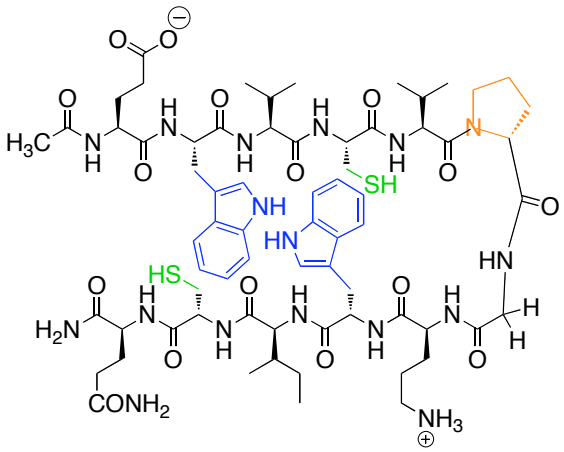


**4** **5**

**Figure 5.9.** β-hairpin peptides **4** and **5** where a $^{D}$Pro-Gly turn (orange) has been implemented in place of the Asn-Gly turn of **1b** and **2b**.

**iii. Hairpin monomer generation 3: WCWC.** Dithiol hairpin monomers **4** or **5** were each combined in solution with monomers **B** and **C** at varying concentrations and ratios, however the major peptide-containing product in each case was the hairpin cyclized by an intramolecular disulfide. It is likely that upon initiation of disulfide exchange, the intramolecular peptide disulfides formed rapidly due to the close proximity of the cysteine residues in each case, and that the entropic cost to break these disulfide bonds was too great, resulting in a quite limited DCL. Furthermore, attempts to diversify the library via templation did not succeed in shifting the library composition, indicating that either there was no recognition of the nucleotides evaluated, or that binding was not strong enough to overcome the entropic cost of reducing either cyclic **4** or **5** to form a more effective receptor.

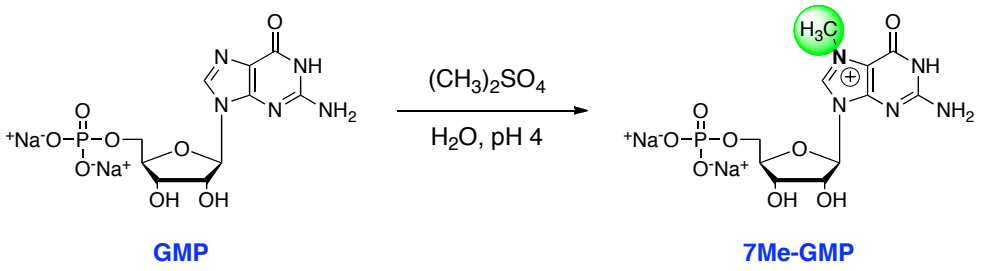
The hairpins were thus redesigned with the Cys residues further apart from each other in attempt to generate a more diverse, useful DCL for future templation studies. A single β-hairpin was synthesized with diagonal Cys residues in positions 4 and 11, in place of the two lysine residues of WKWK (Figure 5.10). While the intrinsic right-handed twist of the β-hairpin brings the diagonal Trps closer together than expected from a flat projection, the diagonal Lys residues are positioned further apart.[23] The preservation of the diagonal Trp residues of WKWK was also expected to aid in nucleotide recognition. The necessity of an acidic residue to achieve total solubility in a DCL with **B** and **C** became increasingly apparent, as significant solubility problems resulted when the N-terminal residue was synthesized as a neutral glutamine in attempt to maintain the same overall peptide charge as **4** and **5**.



**6**

**Figure 5.10.** Structure of hairpin monomer **6** with diagonal Trp residues (blue), and two cysteine residues (green) in place of the Lys residues of WKWK. The peptide also contains an N-terminal Glu for solubility and a $^{D}$Pro-Gly turn (orange).

**iv. Nucleotide templation studies.** To further investigate the recognition of nucleotides, DCC screens were performed with the optimized β-hairpin monomer **6** and the small molecule dithiols **B** and **C**. Efforts were focused on the screening of guanosine 5'-monophosphate (GMP), largely due to the known higher affinity of WKWK for guanosine bases. In addition, this provides an appropriate comparison in evaluating the selective recognition of the methylated analogue, 7-methyl GMP, as the forced positive charge can be expected to enhance recognition. 7-methyl GMP was prepared directly from GMP with dimethyl sulfate (Scheme 5.2).[24] The monophosphate nucleotides were primarily investigated as opposed to the triphosphate nucleotides to minimize any potential charge repulsion between the phosphate tails and the negatively charged aromatic monomers, however this is not expected to drastically influence the results of the DCL screens.




**Scheme 5.2.** Synthesis of 7Me-GMP.

Preliminary dynamic combinatorial libraries were prepared by mixing equimolar amounts of dithiols **6**, **B**, and **C** (0.3 mM each) in basic solution (Figure 5.11). Each small molecule dithiol building block was dispersed in water and dissolved upon addition of NaOH, while a stock solution of **6** was prepared in water. The three building blocks were combined to achieve the desired concentrations, and in two of the three DCLs, either GMP or 7Me-GMP was added at a concentration of 0.3 mM. The pH of each reaction was then adjusted to approximately 8 with NaOH, and the libraries were allowed to oxidized in capped HPLC vials at room temperature. All libraries remained completely soluble with no visible precipitation. The library compositions were analyzed by analytical HPLC daily, and chromatographic separation was optimized for the hairpin-containing macrocycles.

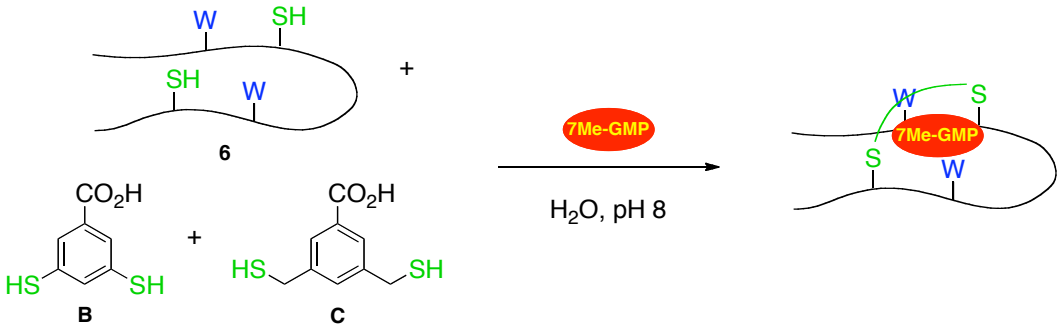



**Figure 5.11.** Dynamic combinatorial library screens containing dithiol monomers **6**, **B**, and **C**, templated with both GMP and 7Me-GMP.

After four days of equilibration, the amplification of two species was observed in the templated libraries, both of which were amplified more significantly in the presence of 7Me-GMP (Figure 5.12). Purification of the two peaks at 8 and 12.5 minutes followed by direct infusion mass spectrometry analysis, led to the identification of these two macrocycles as **6B** and $\mathbf{6B_2}$, respectively (Figure 5.13). While **6B** clearly exists in a higher concentration than $\mathbf{6B_2}$ in all cases, the extent of amplification of $\mathbf{6B_2}$ appears to be greater due to the lower thermodynamic stability of this macrocycle in the untemplated reaction. The apparent enhanced selectivity of both macrocycles for 7Me-GMP is likely due to a combination of cation-π interactions and additional hydrophobic interactions with the two Trp residues and/or the aromatic ring(s) of **B**. It is evident that the diagonal placement of the cysteine residues in **6** facilitates greater library diversity than with **4** or **5**, reducing the extent of intramolecular disulfide formation.

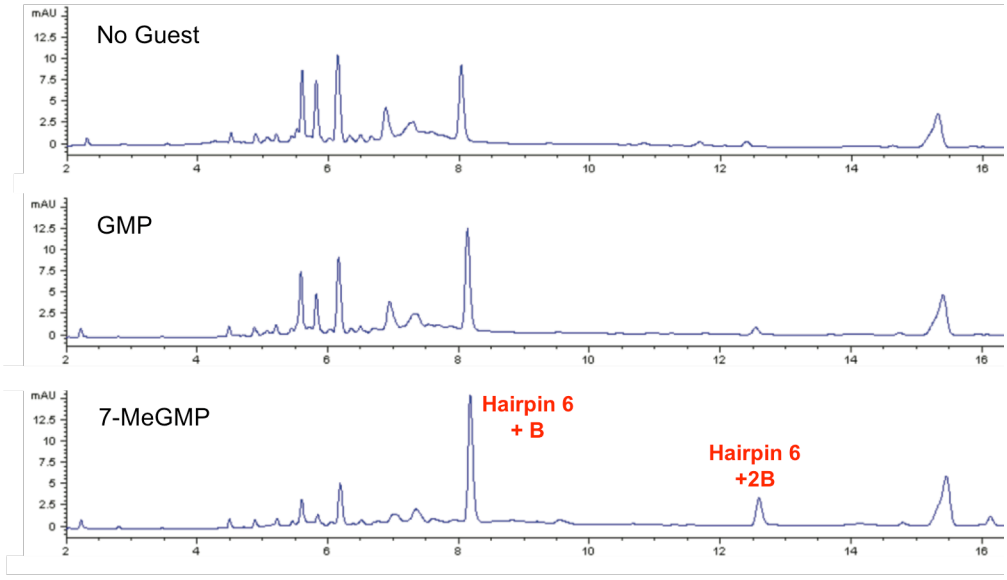



**Figure 5.12.** The analytical HPLC traces at 280 nm of DCLs containing monomers **6**, **B**, and **C** (0.3 mM each), untemplated (top) and in the presence of GMP (0.3 mM, middle) and 7Me-GMP (0.3 mM, bottom) after 4 days of equilibration.

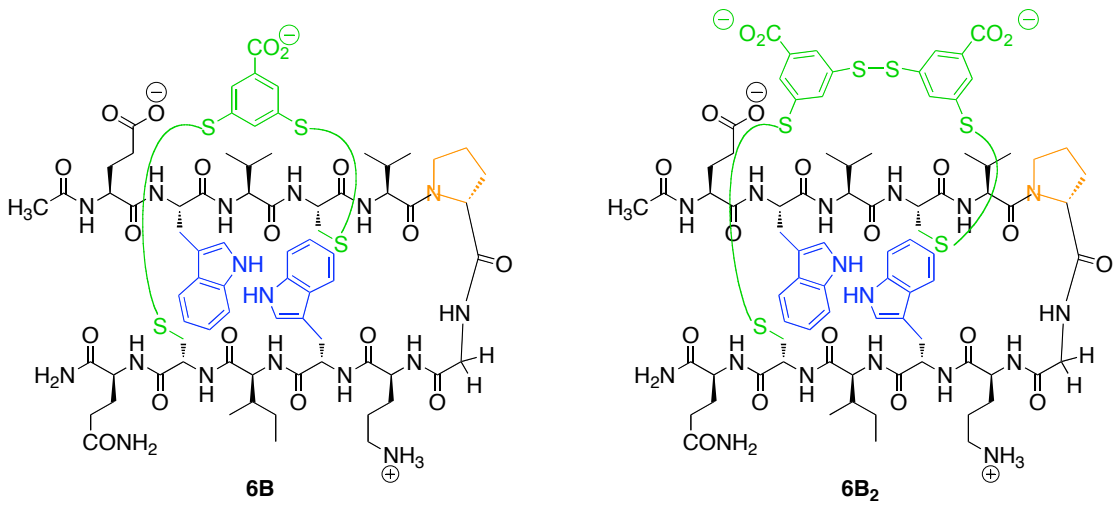



**Figure 5.13.** Cyclic hairpin structures amplified most significantly in the presence of 7Me-GMP.

To further pursue this recognition of 7Me-GMP by the identified macrocycles, subsequent screens were evaluated containing only monomers **6** and **B**, as monomer **C** was not included in either of the amplified macrocycles. In the first screen, the same monomer concentrations were used as in the previous DCLs (0.3 mM each), however the nucleotide concentration was increased to 0.6 mM. The DCLs were investigated by LC-MS in this case to allow for a better understanding of the composition of the entire library, and they were monitored for longer equilibration times. After two days, a 150% increase in the concentration of the **6B** macrocycle was observed in the presence of GMP (as determined by the change in the absolute peak area), in comparison to a 300% increase in the 7Me-GMP templated library (Figure 5.14). The amplification of the **6B$_2$** receptor was slightly greater in each case, resulting in a 200% and 430% change via GMP and 7Me-GMP recognition, respectively. This again illustrates the selectivity for the cationic 7Me-GMP nucleotide. The observed amplification was also found to be highly concentration dependant, as a similar degree of amplification was not observed in a

second screen containing both monomers at higher concentrations (1 mM each). This concentration dependence is not unprecedented in DCC libraries.

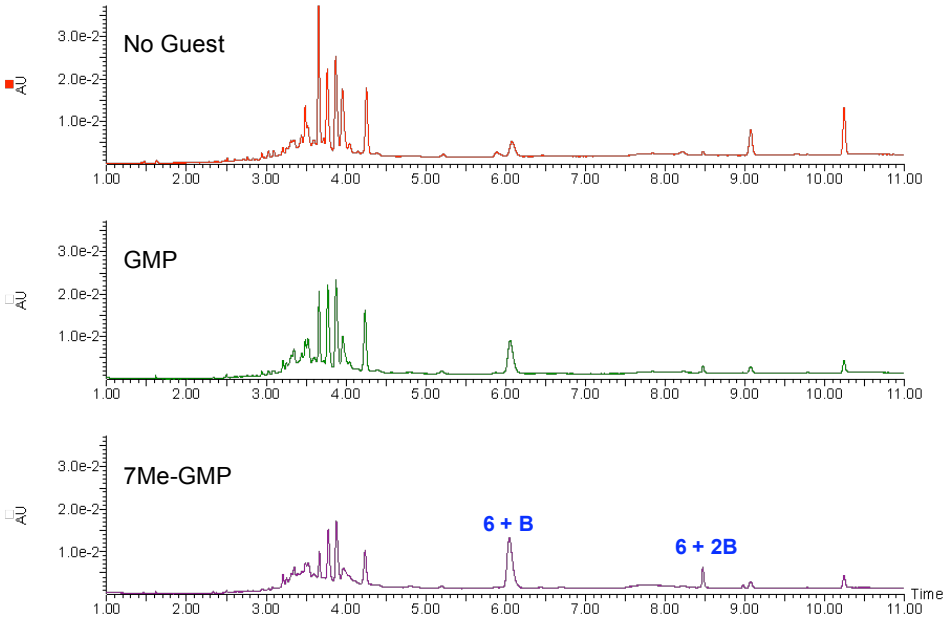



**Figure 5.14.** Part of the analytical HPLC traces at 280 nm of DCLs consisting of monomers **6** and **B** (0.3 mM each), untemplated (red) and in the presence of GMP (0.6 mM, green) and 7Me-GMP (0.6 mM, purple) after 2 days of equilibration.

Analysis of the DCLs at longer time points did not reveal library equilibration as expected, but instead a significant amount of decomposition, preventing a full evaluation of the extent of amplification, and therefore the degree of recognition. This is possibly due to the tendency of peptides to degrade in solution over time, particularly in basic media. It is feasible to envision that the peptide backbone bonds get cleaved over time, resulting in a complete breakdown of the libraries, while other side reactions such as the oxidation of the indole rings of Trp are also feasible. While the basic conditions required for disulfide exchange do not appear to be ideal for the equilibration of hairpin peptide monomers, the initial amplification of **6B** and **6B$_2$** is still promising in the identification of novel peptide receptors for nucleotides

**v. Stability of 7-methyl GMP.** In addition, the amplification of both **6B** and **$6B_2$** in the 7Me-GMP templated library appeared to be short-lived, which prompted an investigation of the stability of the nucleotide templates themselves under the conditions required for equilibration. Although such fleeting peak enhancement is often the result of kinetic amplification prior to reaching the overall lowest energy state of the library, in this case it appeared to be due to the instability of the 7Me-GMP template under the equilibration conditions. From the first time point through day 6 of equilibration of the DCL, little change in the size of the HPLC peak corresponding to GMP was observed, which suggests the nucleotide is stable throughout this analysis (Figure 5.15). In contrast, the HPLC peak corresponding to the 7Me-GMP template decreased substantially after only one day and was entirely gone after 6 days (Figure 5.16). The rate of degradation also appeared to be pH dependant, increasing with an increase in pH, which is precedented in the literature. As the concentration of 7Me-GMP decreased, it is likely that the formation of **6B** and **$6B_2$** was no longer energetically favorable, causing re-equilibration of the library and thus a decrease in the concentration of the receptors. Since the 7Me-GMP decomposition begins well before equilibrium of the DCL is reached, both the level of affinity and extent of selectivity for 7Me-GMP over GMP are difficult to fully ascertain.

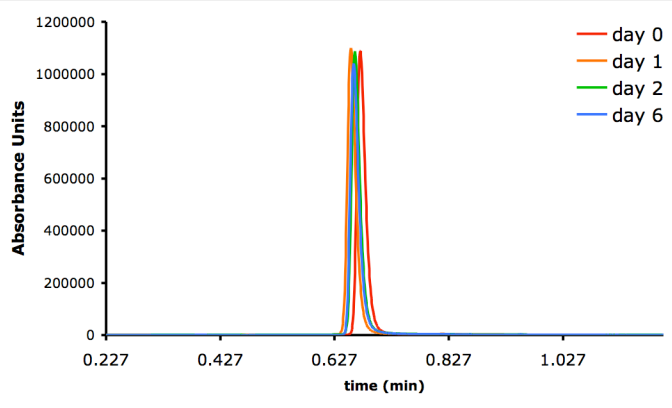



**Figure 5.15.** The GMP template peak in the analytical HPLC traces of the GMP templated library at time 0 (red) and after 1 day (orange), 2 days (green), and six days (blue).

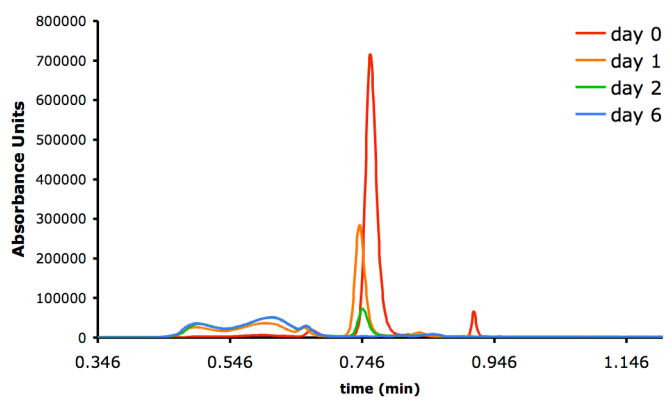
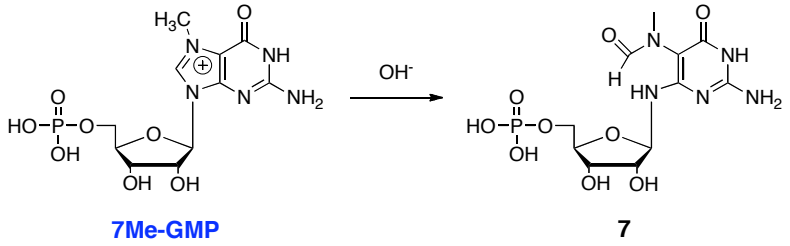



**Figure 5.16.** The 7Me-GMP template peak in the analytical HPLC traces of the 7Me-GMP templated library at time 0 (red) and after 1 day (orange), 2 days (green), and six days (blue).

The source of this apparent instability of 7Me-GMP is believed to be due to the basic solution required for disulfide exchange as opposed to the presence of thiolate anions, as indicated by various control experiments. The decomposition is likely initiated by the irreversible nucleophilic attack of a hydroxide ion on the CH adjacent to the methylated nitrogen to give the ring-opened product **7** (Scheme 5.3).[25] This is a viable mechanism

for the decomposition observed in the DCL, as a new mass 17 Da greater than that expected for 7Me-GMP is observed.




**Scheme 5.3.** Possible pathway of decomposition of 7Me-GMP.

**vi. Isolation of hairpin receptors.** To purse further investigations of the cyclic peptide receptors as potential 7Me-GMP receptors, the preparation and isolation of these molecules was required. This is achieved via DCC in a fairly straightforward manner. A library was prepared with a 1:2 ratio of **6** to **B** and a total concentration of 3 mM. Due to the stability of GMP in solution, the library was templated with GMP (2 mM); however, amplification of **6B** and $\mathbf{6B_2}$ is minimal at these concentrations as described previously. The solution was stirred in a capped vial for 4-5 days before purifying both **6B** and $\mathbf{6B_2}$ and on an analytical HPLC, yielding only a few milligrams or less in each case. The identity and purity of the compounds were confirmed by analytical LC-MS.

**vii. NMR characterization of hairpin receptors.** To investigate the structural differences between the isolated receptors and the linear β-hairpin **6**, NMR was used to characterize the peptides. Initial attempts to dissolve **6B** in a purely aqueous solution proved difficult, but complete solubility of the peptide was achieved in a 40% MeOD solution. Peptide **6** was first investigated under identical conditions, revealing a very well folded β-hairpin. This was demonstrated by a glycine splitting of 0.65 ppm, indicating different chemical environments of the diastereotopic $H_\alpha$ protons, which is

characteristic of a β-hairpin confirmation. Furthermore, distinct shifts of the $H_\alpha$ protons of various residues relative to what would be expected for an unfolded system further confirmed a high overall degree of hairpin structure. While it should also be noted that the degree of β-hairpin folding is known to increase in methanol, this solvent system was used for an accurate comparison to the cyclic receptors.

In contrast, the degree of folding of **6B** should be less affected in methanol as a result of the preformed covalent tether linking the two strands of the peptide. However, both 1D and 2D TOSCY experiments of **6B** revealed a substantially less hairpin-like structure in 40% MeOD. Little to no splitting of the glycine protons was observed, while the upfield shifting of the $H_\alpha$ protons of various residues such as Ile, Val, and Orn further confirmed the non-hairpin like structure. It is clear that by bridging the two strands at the diagonal cysteine residues via a single **B** monomer, the peptide structure is significantly distorted from that of **6**. Due to this apparent structural change, it is likely that the binding of both nucleotides by the **6B** receptor is quite different than the binding of GMP by WKWK. Although the structure of **6B** in the presence of GMP or 7Me-GMP was not evaluated, little change in folding is expected due to the fixed nature of the cyclic peptide. While a comparison to $\mathbf{6B_2}$ would have been interesting to determine if hairpin structure can be restored with a longer bridge between the two peptide strands, this receptor proved to be substantially less soluble under these conditions, resulting in concentrations of peptide in solution that were too low for detection by NMR.

## C. Conclusion

In conclusion, this study illustrates the use of structured β-hairpin peptides in the context of a dynamic combinatorial library, for the identification of selective peptide receptors for GMP and 7Me-GMP. Peptide **6** was designed based on the known nucleotide hairpin receptor, WKWK, with two cysteine residues to allow for disulfide exchange of the peptide with the small aromatic dithiol monomers, **B** and **C**. However, obtaining complete solubility with such a system containing monomers with quite different structures, polarities, and hydrophobicities is not trivial.

Library screens in the presence of GMP and 7Me-GMP revealed the amplification of two cyclic structures, **6B** and $\mathbf{6B_2}$, each of which is modestly amplified in the presence of GMP, and more significantly in the presence of 7Me-GMP. It appears that by covalently bridging the two peptide strands of **6** and potentially forming a tighter binding pocket, more favorable interactions with the nucleotide templates are created. Furthermore, the methyl group on the guanosine base serves to enhance these interactions, either through additional hydrophobic contacts and/or through favorable cation-π interactions with the Trp side chains of **6** or the aromatic rings of **B**.

Interestingly, the binding of both GMP and 7Me-GMP to the identified peptide macrocycles appears to compete with the folding of the linear peptide (Figure 5.17). While **6** maintains a well-folded, hairpin-like structure, the structure of **6B** is quite different, no longer resembling a β-sheet. It can be concluded that in the presence of both GMP and 7Me-GMP, **6B** exists in a higher concentration relative to the untemplated reaction due to a greater energy gain as a result of the binding of $\mathbf{6B_2}$ versus the folding of **6**. While the stability of both the library and the 7Me-GMP template proved to be non-

ideal under the conditions required for disulfide exchange, this DCC system still proved to be effective in demonstrating novel cyclic peptide receptor structures. Such receptors selective for 7Me-GMP may provide greater insight into native nucleotide-protein interactions, as well as aid in elucidating the role of such interactions in gene expression.

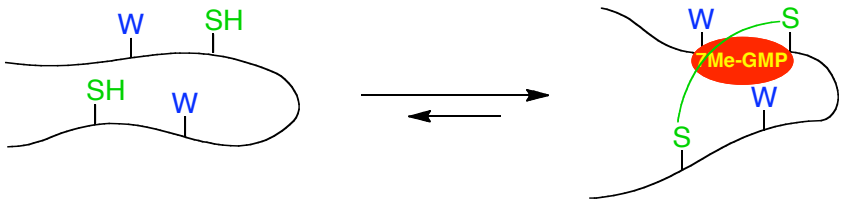



**Figure 5.17.** The competing equilibria: folding of hairpin **6** and binding of macrocycles **6B** or **6B**$_2$.

## D. Experimental Section

**i. Peptide synthesis.** All peptide synthesis was performed on an automated Applied Biosystems Pioneer Peptide Synthesizer using a PEG-PAL-PS resin. Peptides were synthesized on a 0.1 mmol scale by standard Fmoc solid phase synthesis. All amino acids with functionality were protected during synthesis. Coupling reagents were HOBt/HBTU in DMF. Fmoc deprotections were carried out in a solution of 2% 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) and 2% piperidine in DMF and, the N-terminus was acylated with a solution of 5% acetic anhydride and 6% 2,6-lutidine in DMF. Cleavage of the peptide from the resin and of the side chain protecting groups was performed by hand with a cocktail of 95% trifluoroacetic acid (TFA)/2.5% ethanedithiol (EDT)/1.0% triisopropylsilane/2.5% $H_2O$ for all Cys containing peptides, and 95% trifluoroacetic acid (TFA)/2.5% triisopropylsilane/2.5% $H_2O$ for all other peptides for 3 hours. TFA was evaporated and the product was precipitated with cold ether. The water-soluble peptides were extracted with water and lyophilized. Peptides were purified by

semipreparative reversed-phase HPLC on a C18 column at a flow rate of 4 mL/min. Peptides were purified with a linear gradient of A and B (A: 95% $H_2O$/5% $CH_3CN$ with 0.1% TFA, B: 95% $CH_3CN$/5% $H_2O$ with 0.1% TFA) and elution was monitored at 214 and 280 nm. Once purified, peptides were lyophilized to powder and characterized by +ESI-MS.

**ii. 7-methyl GMP synthesis.** The disodium salt of guanosine 5'-monophosphate (0.2 g, 0.49 mmol) was dissolved in 1.6 mL of water rand the pH was adjusted to approximately 4.5 with 1 M HCl. As the pH was lowered, additional volume of water was required to maintain adequate solubility. Dimethyl sulfate (292 mL, 3.1 mmol) was added dropwise via syringe pump over a period of 1 hour at room temperature with continuous stirring, and then the reaction was allowed to stir for an additional 3 hours. During the course of the reaction the pH was maintained between approximately 3.0 and 4.5 to prevent esterification of the phosphate groups. The reaction was pass through an ion exchange column of Dowex-1 $Cl^-$ (50-100 mesh, 11 × 2 cm) and eluted with $H_2O$. The product eluted quickly and 14 fractions were collected, the first 7 of which showed a major ultraviolet fluorescent spot. Fractions 1 – 7 were pooled and the product was lyophilized to powder, to obtain pure product. Further purification via re-precipitation of the product with the addition of ethanol has been reported, however this not found to be necessary.[19] $^1$H NMR ($D_2O$, 300 MHz): δ = 8.97 (s, 1H), 5.90 (d, 1H), 4.48 (t, 1H), 4.28 (t, 1H), 4.22 (m, 1H), 4.06 (m, 1H), 3.94 (m, 1H), 3.92 (s, 3H).

**iii. Dynamic combinatorial library screens.** Building block **B** and **C** stock solutions were each prepared in water, adding sufficient 1.0 M aqueous NaOH to fully dissolve the monomers, using sonication when necessary. Peptide stocks were prepared

in water and the concentrations were determined by UV-Vis in 5M guanidine hydrochloride, using the absorbance of the tryptophan residues at 280 nm ($\varepsilon$ = 5,690 $M^{-1}$ $cm^{-1}$ per Trp).[26] Similarly, nucleotide solutions were prepared in water and their concentrations were determined by UV-Vis (GMP: $\varepsilon$ at 253 nm = 13,700 $M^{-1}$ $cm^{-1}$; 7Me-GMP: $\varepsilon$ at 297 nm = 3,537 $M^{-1}$ $cm^{-1}$). The relevant building blocks were combined to achieve a 1 mL solution of the desired concentrations (300 μM each, unless indicated otherwise), and when necessary, aliquots of nucleotides guests were added to the reactions to reach the desired final concentration. Any remaining volume was made up with water, and the pH of each solution was adjusted to approximately 8 using 1.0 M NaOH. The vials were capped and analyzed at various time points, and the pH of each solution was monitored over time.

**iv. Analytical HPLC and LC/MS.** The analytical HPLC was carried out daily (until significant library decomposition was observed) on an Agilent Rapid Resolution LC system, equipped with an online degasser, binary bump, autosampler, heated column compartment, and diode array detector. HPLC separations were performed using a water/acetonitrile solvent system (A: 95% $H_2O$/5% $CH_3CN$ with 0.1% TFA, B: 95% $CH_3CN$/5% $H_2O$ with 0.1% TFA; gradient: 0-30% B, 0-4 minutes / 30-35% B, 4-12 minutes / 35-45% B, 12-20 minutes) with a column temperature of 55 °C, a flow rate of 2.0 mL/min, and 5 μL injection volumes. Reactions were monitored at 214, 280, and 320 nm. LC-MSD was performed on a Waters Acquity LC-MSD system with an in-line single quadrupole mass spectrometer. LC-MS separations were performed using a water/acetonitrile solvent system (A: 95% $H_2O$/5% $CH_3CN$ with 0.2% formic acid, B: 95% $CH_3CN$/5% $H_2O$ with 0.2% formic acid; gradient: 0-25% B, 0-2 minutes / 25-30%

B, 2-3 minutes / isocratic at 30% B, 3-5 minutes / 30-32% B, 5-7 minutes / 32-75% B, 7-13 minutes) with a column temperature of 55 °C, a flow rate of 0.6 mL/min, and 2 μL injection volumes. Positive electrospray ionization was used to determine the peak masses, however it should be noted that the masses of macrocycles consisting of only **B** and **C** could not be identified via positive ionization. The peak areas were integrated at 280 nm.

**v. Synthesis and isolation of hairpin receptors.** Libraries were prepared as described above on a 1-3 mL scale containing 1 mM **6** and 2 mM **B** and templated with GMP (2 mM), despite the lack of significant amplification at these millimolar concentrations. After approximately four days the libraries were purified on a Waters Alliance analytical HPLC with an Atlantis C18 column (4.6 × 150 mm, 5 μm). The libraries were chromatographed (50 – 100 μL automated injections) using standard water/acetonitrile peptide synthesis mobile phases (A: 95% $H_2O$/5% $CH_3CN$ with 0.1% TFA, B: 95% $CH_3CN$/5% $H_2O$ with 0.1% TFA; gradient: 0-25% B, 0-8 minutes / 25-30% B, 8-12 minutes / isocratic at 30% B, 12-20 minutes / 30-32% B, 20-28 minutes / 32-57% B, 28-42 minutes) with a column temperature of 50 °C and a flow rate of 1.0 mL/min. Reactions were monitored at 280 nm (Figure 5.18). Purified **6B** and **$6B_2$** were analyzed for purity by +ESI (Figure 5.19) and lyophilized to powder.

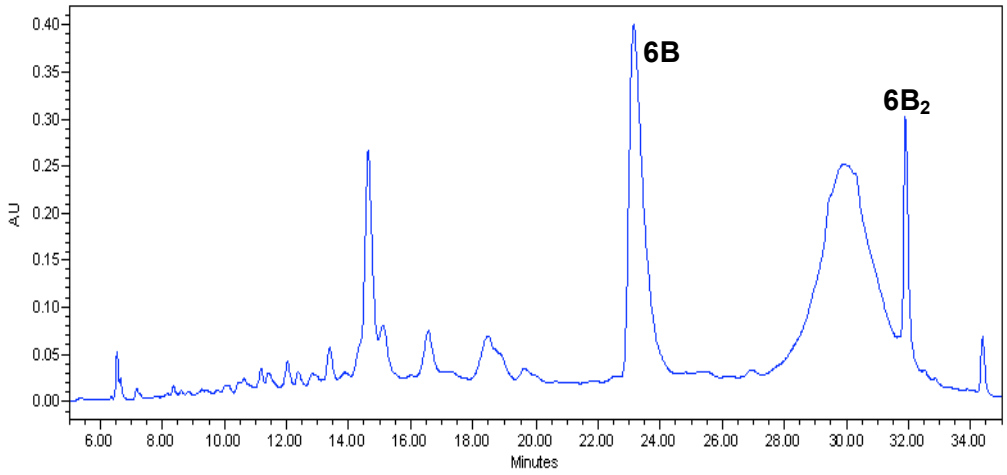



**Figure 5.18.** Purification of peptides **6B** and **$6B_2$** on an analytical HPLC (280 nm).

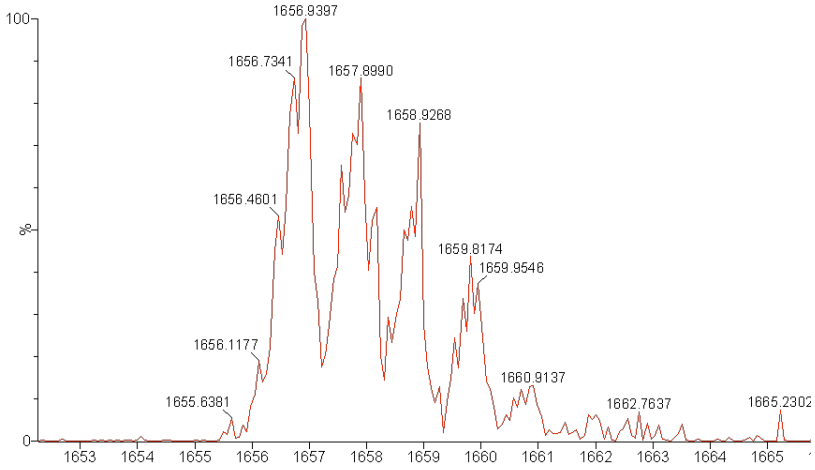



**Figure 5.19.** Representative mass spectra: purified **6B** (+ESI, $[M+H]^+$ = 1656.9)

**vi. NMR spectroscopy.** Peptide NMR samples were analyzed on a Varian Inova 600-MHz instrument at 25 °C. 1D NMR spectra were collected using 32K data points and 32 scans using a 1-3 second presaturation. All 2D NMR experiments used pulse sequences from the Chempack software. 2D TOSCY experiments were taken with 32 scans in the first dimension and 128 scans in the second dimension. All spectra were analyzed using standard window functions (sinbell and Gaussian with shifting).

Presaturation was used to suppress the water resonance. Peptide proton assignments were made following standard methods.[27]

To investigate deamidation of the hairpin turn, peptide **3** was dissolved in 50% $H_2O$ and 50% $D_2O$ and analyzed by 1D NMR (Figure 5.20). A second sample of **3** was dissolved in $H_2O$ at pH 8-9 and allowed to stir for 3 days. The peptide was purified by HPLC and then analyzed similarly by 1D NMR (Figure 5.21). The signals corresponding to a single Asn NH proton and a single Gln NH proton were identified and integrated in each case (Figure 5.22).

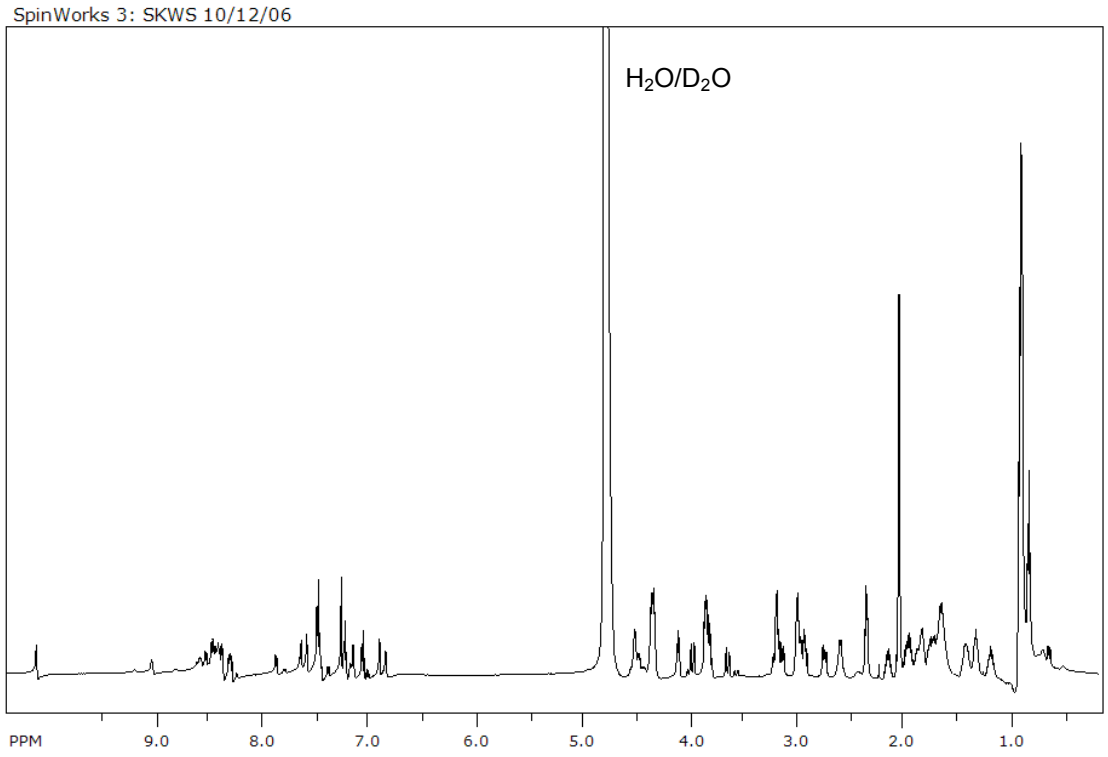



**Figure 5.20.** Full $^1$H NMR of **3** at 25 °C.

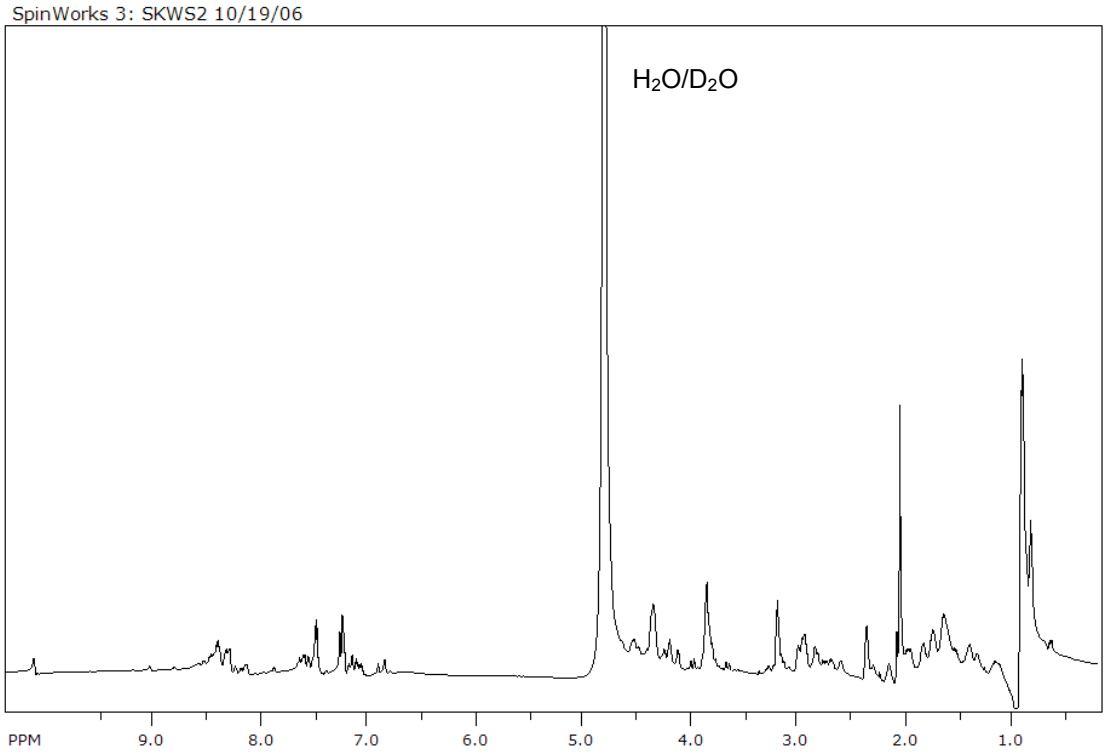



**Figure 5.21.** Full $^{1}$H NMR of **3** at 25 °C after exposure to basic solution for 3 days, resulting in a mixture of intact peptide and peptide that has undergone deamidation.

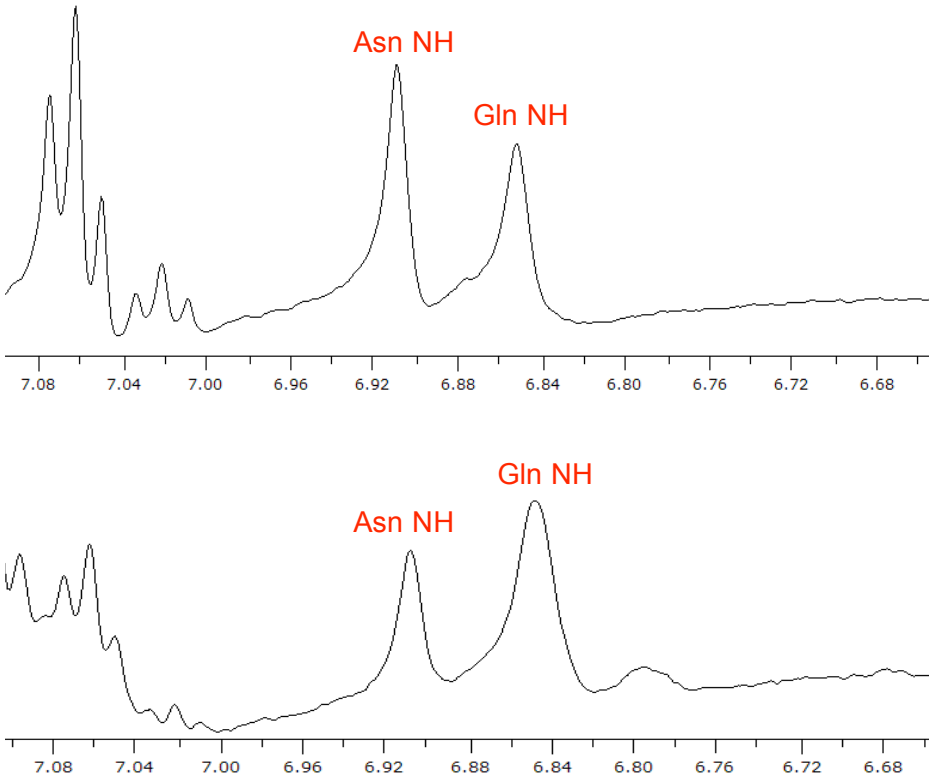



**Figure 5.22.** Asn and Gln NH $^{1}$H NMR signals of **3** at 25 °C before (top) and after (bottom) exposure to basic solution.

Samples of **6** and **6B** were prepared in 40% MeOD and 60% $D_2O$ with 0.5 mM DSS. 1D spectra were taken of each, as well as 2D TOSCY spectra allowing for the assignment and comparison of amino acid $H_\alpha$ protons (Table 2.6).

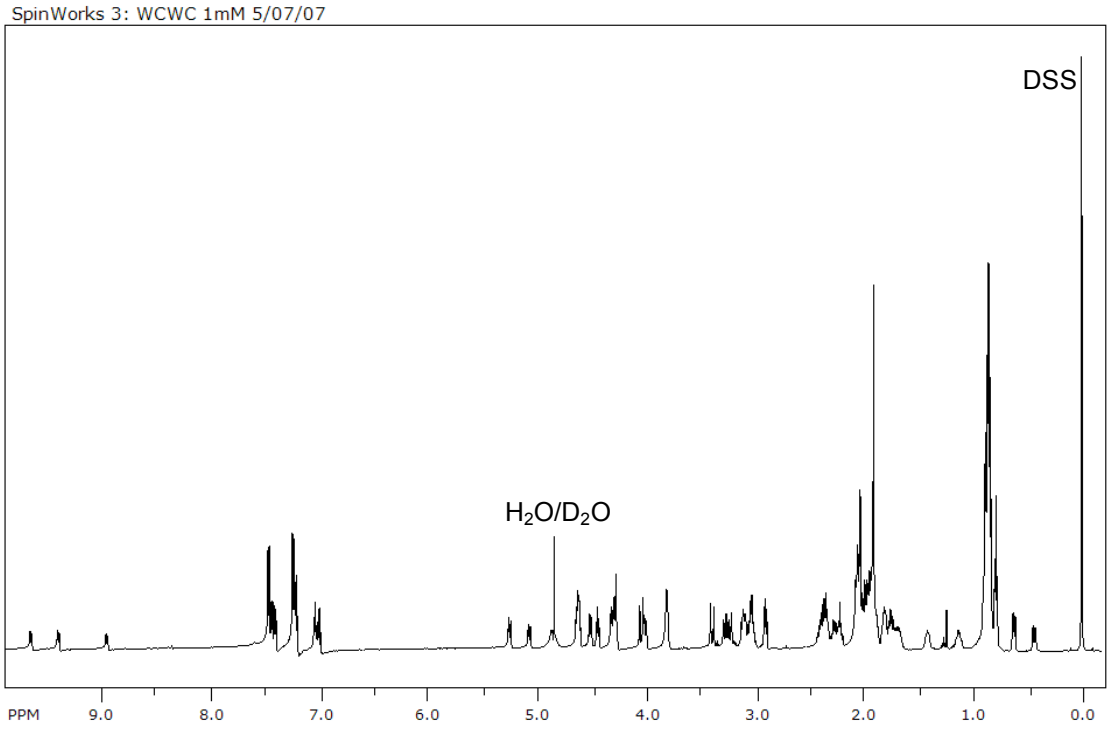



**Figure 5.23.** $^{1}$H NMR of **6** at 25 °C.

**Table 5.2.** Proton $H_\alpha$ chemical shift assignments for **6** ($EW_1V_1C_1V_2{}^{D}PGOW_2IC_2Q$) and **6B** at 25 °C.

| Residue | **6** $H_\alpha$ | **6** + **B** $H_\alpha$ |
|---|---|---|
| **Glu** | 4.35 or 4.42 | 4.15 or 4.28 |
| **Trp$_1$** | 5.19 or 5.32 | 4.67 or 4.80 |
| **Val$_1$** | 4.56 | 4.28 |
| **Cys$_1$** | 4.42 | 4.63 or 5.37 |
| **Val$_2$** | 4.09 | 4.28 |
| **$^{D}$Pro** | 4.28 or 3.83 | 4.42 |
| **Gly** | 4.02 and 3.37 | 3.80 |
| **Orn** | 4.66 | 3.94 |
| **Trp$_2$** | 5.19 or 5.32 | 4.67 or 4.80 |
| **Ile** | 4.64 | 4.12 |
| **Cys$_2$** | 4.42 | 4.63 or 5.37 |
| **Gln** | 4.35 or 4.42 | 4.15 or 4.28 |

[1] Gellman, S. H. *Curr. Opin. Struct. Biol.* **1998**, *2*, 717-725.

[2] Griffiths-Jones, S. R.; Maynard, A. J.; Sharman, G. J.; Searle, M. S. *Chem. Commun.* **1998**, 789-790.

[3] (a) Ramierz-Alvarado, M.; Kortemme, T.; Blanco, F. J.; Serrano, L. *Bioorg. Med. Chem.* **1999**, *7*, 93-103. (b) Stotz, C. E.; Topp, E. M. *J. Pharm. Sci.* **2004**, *93*, 2881-2894. (c)

[4] (a) Searle, M. S.; Griffiths-Jones, S. R.; Skinner-Smith, H. *J. Am. Chem. Soc.* **1999**, *121*, 11615-11620. (b) Kiehna, S. E.; Waters, M. L. *Protein Sci.* **2003**, *12*, 2657-2667.

[5] Tatko, C. D.; Waters, M. L. *J. Am. Chem. Soc.* **2002**, *124*, 9372-9373.

[6] Tatko, C. D.; Waters, M. L. *Protein Sci.* **2003**, *12*, 2443-2452.

[7] Moodie, S. L.; Mitchell, J. B. O.; Thorton, J. M. *J. Mol. Biol.* **1996**, *263*, 486-500.

[8] (a) Butterfield, S. M.; Waters, M. L. *J. Am. Chem. Soc.* **2003**, *125*, 9580-9581. (b) Butterfield, S. M.; Sweeney, M. M.; Waters, M. L. *J. Org. Chem.* **2005**, *70*, 1105-1114.

[9] Lindahl, T. *Nature*, **1981**, *290*, 363-364.

[10] Cowling, V. H. *Biochem J.* **2010**, *425*, 295-302.

[11] Zhang, Z.; Lonnberg, H.; Mikkola, S. *Org. Biomol. Chem.* **2003**, *1*, 3404-3409.

[12] Richter, J. D.; Sonenberg, N. *Nature*, **2005**, *433*, 477-480.

[13] De Benedetti, A.; Graff, J. R. *Oncogene* **2004**, *23*, 3189.

[14] Natarajan, A.; Moerke, N.; Fan, Y.-H.; Chen, H.; Christ, W. J.; Wagner, G.; Halperin, J. A. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2657-2660.

[15] (a) Kamiichi, K.; Doi, M.; Nabae, M.; Ishida, T.; Inoue, M. *J. Chem. Soc. Perkin Trans. II* **1987**, 1739-1745. (b) Stolarski, R.; Sitek, A.; Stepinski, J.; Jankowska, M.; Oksman, P.; Temeriusz, A.; Darzynkiewicz, E.; Lonnberg, H.; Shugar, D. *Biochim. Biophys. Acta* **1996**, *1293*, 97-105.

[16] Ruszczynska, K.; Kamienska-Trela, K.; Wojcik, J.; Stepinski, J.; Darzynkiewicz, E.; Stolarski, R. *Biophysical Journal*, **2003**, *85*, 1450-1456.

[17] Gallivan, J. P.; Dougherty, D. A. *Proc. Natl. Acad. Sci.*, **1999**, *96*, 9459-9464.

[18] Niedzwiecka, A.; Marcotrigiano, J.; Stepinski, J.; Jankowska-Anyszka, M.; Wyslouch-Cieszynska, A.; Dadlez, M.; Gingras, A.-C.; Mak, P.; Darzynkiewicz, E.; Sonenberg, N.; Burley, S. K.; Stolarski, R. *J. Mol. Biol.* **2002**, *319*, 615-635.

[19] (a) Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *J. Am. Chem. Soc.* **2000**, *122*, 12063-12064. (b) Ramström, O.; Lehn, J.-M. *ChemBioChem* **2000**, *1*, 41-48.

[20] Stotz, C. E.; Borchardt, R. T.; Middaugh, C. R.; Siahaan, T. J.; Vander Velde, D.; Topp, E. M. *J. Peptide Res.* **2004**, *63*, 371-382.

[21] Patel, K.; Borchardt, R. T. *Pharm. Res.* **1990**, *7*, 703-711.

[22] Stanger, H. E.; Gellman, S. H. *J. Am. Chem. Soc.* **1998**, *120*, 4236-4237.

[23] Syud, F. A.; Stanger, H. E.; Gellman, S. H. *J. Am. Chem. Soc.* **2001**, *123*, 8667-8677.

[24] Hendler, S.; Furer, E.; Srinivasan, P. R. *Biochemistry* **1970**, *9*, 4141-4153.

[25] Darzynkiewicz, E.; Stepinski, J.; Tahara, S. M.; Stolarski, R.; Ekiel, I.; Haber, D.; Neuvonen, K.; Lehikonen, P.; Labadi, I.; Lonnberg, H. *Nucleosides Nucleotides* **1990**, *9*, 599-618.

[26] Edelhoch, H. *Biochemistry*, **1967**, *6*, 1948-1954.

[27] Wuthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley-Interscience: New York, 1986.

# CHAPTER VI

# BINDING INDUCED FOLDING OF A PHOTOCONTROLLED β-HAIRPIN PEPTIDE

## A. Background and Significance

**i. β-Hairpin model systems for the recognition of nucleotides.** Short peptide sequences that adopt a specific secondary structure are often used as model systems to mimic a certain protein fold and study the factors that contribute to protein folding and protein recognition. Motivated by the prevalence of β-sheet motifs throughout protein structures and the desire to better understand the factors contributing to their stability, short, monomeric β-hairpin peptides have been studied as models for antiparallel β-sheets.[1] These β-sheet mimics contain two antiparallel β-strands connected by a turn sequence, such as an Asn-Gly type I' turn motif, which nucleates the hairpin structure.[2]

As discussed previously in chapter five, it has been found that β-hairpin peptides can serve as novel molecular receptors for biologically relevant aromatic compounds, such as nucleotides. Specifically, the WKWK hairpin (Figure 5.1) has been shown to be a quite effective receptor for adenosine 5'-triphosphate (ATP), as the diagonal tryptophan-tryptophan pair in the non-hydrogen bonding sites of the hairpin provides a binding cleft for aromatic intercalation, while the lysine side chains provide favorable electrostatic interactions with the phosphate groups.[3] ATP was found to bind the β-hairpin with a dissociation constant of 170 μM, and a total binding energy of approximately -5 kcal/mol.

**ii. Photoswitchable β-hairpins.** The turn sequence of the β-hairpin has been found to be crucial in ensuring proper nucleation of the β-sheet structure. While WKWK contains an Asn-Gly turn, it has also been shown that the D-Pro-Gly segment is a very strong promoter of β-hairpin formation, and superior to the L-Asn-Gly segment in this regard.[4] The utility of the D-Pro-Gly turn sequence for fundamental β-sheet analysis is particularly enhanced due to the fact that swapping L-proline for D-proline destroys hairpin structure, as L-proline negates formation of β-hairpin loops. While it is useful to be able to prepare and study β-hairpin sequences in both the folded and unfolded state based on the implementation of either a D-Pro-Gly or L-Pro-Gly turn respectively, it would be equally interesting to study a structure that is able to easily convert between the folded and unfolded state in response to an external stimuli.

The use of light to initiate the reversible conformational control of β-hairpin formation is one way in which this can be achieved. Azobenzene moieties are ideally suited for this application as they give rise to large changes in the distance between moieties appended to the aryl rings upon interconversion of the trans and thermodynamically less favored *cis*-azobenzene isomers. While irradiation at the wavelength of the $\pi \rightarrow \pi^*$ transition converts the trans isomer to the corresponding cis isomer, the reverse process can also be facilitated either thermally or by irradiation of the wavelength of the $n \rightarrow \pi^*$ transition.[5] Incorporating azobenzene photoswitches into small protein motifs has been demonstrated in the amino acid side chains,[6] in the backbone,[7] and even as a β-turn mimic when in its cis form.[8] Specifically, monomeric soluble β-hairpin peptides whose structure is reversibly photocontrolled by an azobenzene-based chromophore that acts as a turn mimetic at the tip of the hairpin have been presented.[9]

Such azobenzene moieties have provided excellent conformational control while allowing for time-resolved folding studies of β-sheet structures with unprecedented temporal resolution.

In one case a photocontrolled hairpin peptide has been demonstrated in the context of a tryptophan zipper-like β-hairpin.[9b] While the original sequence is fully folded at room temperature,[10] substitution of the Asn-Gly turn residue by an azobenzene chromophore considerably destabilizes the hairpin, resulting in an estimated 50% β-hairpin content at 5 °C for the *cis*-azo photostationary state as determined by CD. In comparison, at 30 °C the extent of folding of the *cis*-isomer decreases to 25%, and a two-state transition with a melting temperature of 25 °C has been proposed. In both cases, the hairpin fold is significantly diminished upon isomerization to the *trans*-azo isomer. NMR structures revealed close structural agreement between the *cis*-azo peptide and the native TrpZip peptide for the antiparallel β-strands as well as the tryptophan side chains, whereas the *trans*-azo isomer displays a quite different alignment of the two peptide strands (Figure 6.1).

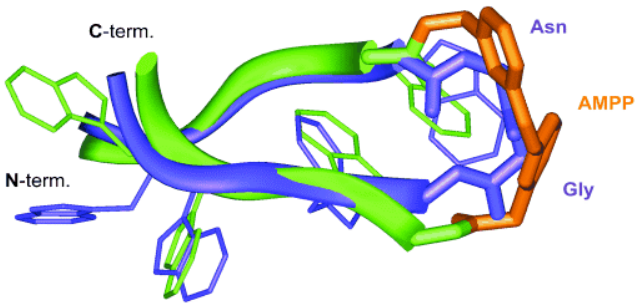



**Figure 6.1.** Backbone superimposition of the lowest energy NMR structure of the *cis*-azo peptide isomer (green ribbons) on the original tryptophan zipper (blue ribbons, PDB code 1LE1). The azobenzene chromophore (orange) and side chains of the tryptophan residues are displayed.

The use of a photoinducible azobenzene turn element to control the folding of β-sheets has also been shown in a 12-residue peptide derived from the GB1 protein.[9a] In its cis configuration the azobenzene moiety mimics the native D-Pro-Gly dipeptide turn; however, it was observed that in this case the *cis*-azo peptide takes on a different fold than the parent peptide. The backbones of the N- and C-strand segments are both flipped by about 180° in the *cis*-azo hairpin, so that the $H_\alpha$ protons which face the opposing strand in one peptide are directed toward the solvent in the other. Although clearly this turn mimic serves to preorganize two peptide strands as a β-hairpin, the more flexible nature of the azobenzene turn favors a different set of intramolecular interactions, leading to significant structural differences from the parent peptide. Furthermore this system was met with extensive aggregation and precipitation of the trans conformation of the hairpin, likely due to intermolecular association of the molecules through hydrophobic interactions and stacking of the azobenzene units, which is not unprecedented.[11]

**iii. Goal of this work.** In this work we set out to investigate the effect of introducing a photoswitchable azobenzene turn moiety into the peptide, WKWK, which is a known receptor for nucleotides. No prior studies have examined a photoinducible β-hairpin in relation to its binding capabilities. We sought to evaluate the impact of the azobenzene turn on the recognition of ATP in both the cis and trans conformation. Furthermore, the correlation between folding and binding was investigated, as well as the potential influence of binding on the ratio of cis and trans peptide isomers. The relationship between binding and hairpin folding is a unique one, and this provides an ideal opportunity to examine both of these elements in a simple system. All prior studies of nucleotide recognition by β-hairpins have made use of well-folded systems, and therefore

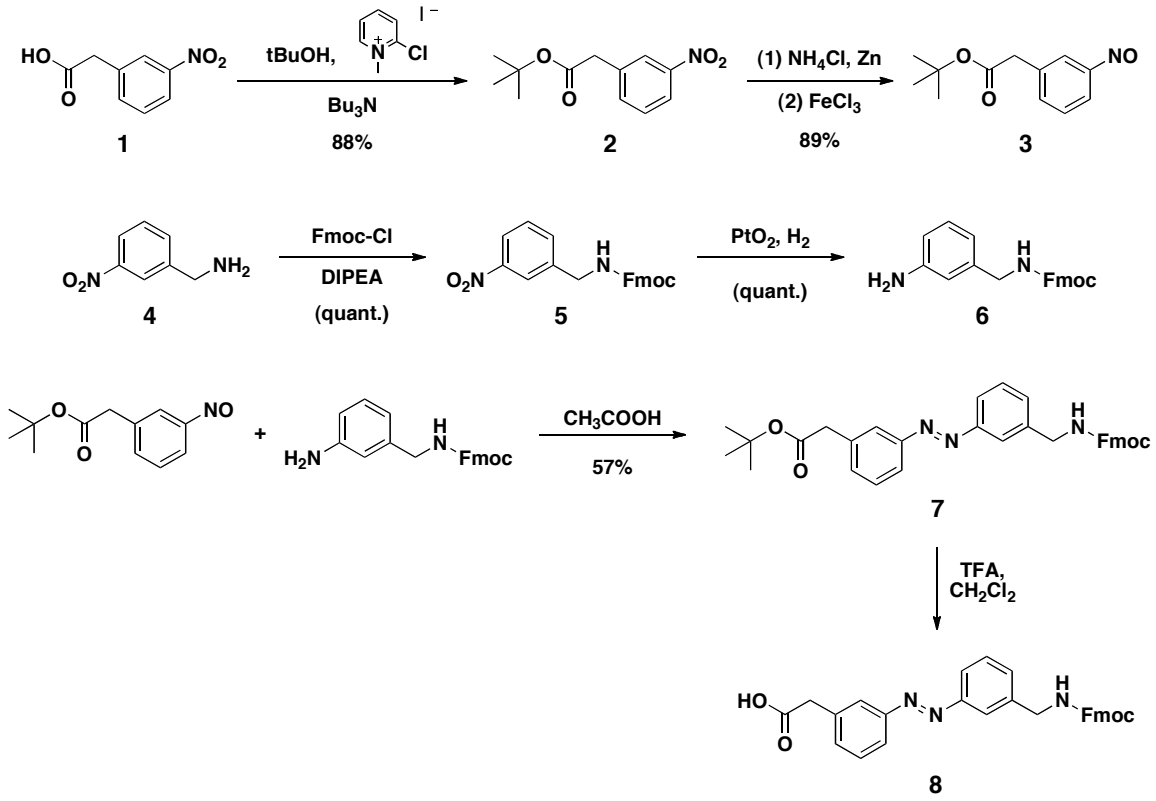
it is not known to what degree of preorganization is required to facilitate binding. Furthermore, the photoswitchable nature of this system provides an opportunity to achieve a greater degree of control over both hairpin folding and recognition through the use of an external light source.

## B. Results and Discussion

**i. Design of a photoswitchable hairpin.** The 12-residue WKWK peptide has been shown to have a high propensity to form a β-hairpin, and has been characterized by NMR to be about 95% folded in an aqueous environment. The type I' Asn-Gly turn segment at positions 6 and 7 nucleates hairpin formation, and was sought to be replaced with an azobenzene-containing turn mimetic. Favorable interactions between the hydrophobic pocket formed by the Trp side chains and the turn element in its cis configuration were expected to favor a β-hairpin structure.

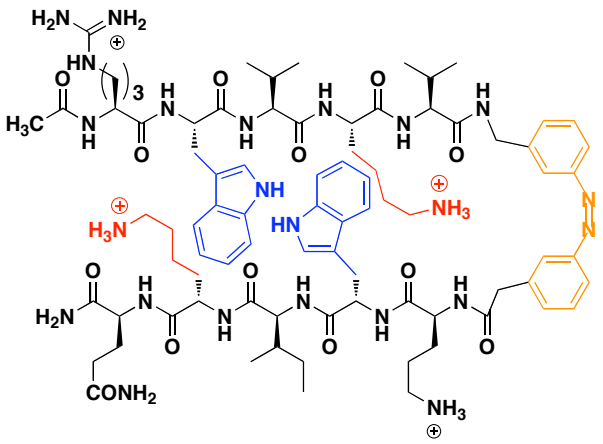
An azobenzene based Fmoc protected amino acid was used for the facile incorporation of this moiety via solid phase synthesis. It has previously been suggested that an amino acid with meta-substituted aryl rings was a more promising turn mimetic than ortho- or para-substituted aryl rings.[12] Similarly, a single methylene group has been found as a more ideal spacer than multiple methylenes between the functional groups and the aryl rings in order to accommodate interstrand interactions between the appended peptides.[12] The synthesis of the Fmoc amino acid **8** relies on the reaction of a nitrosobenzene **3** with an aniline **6** (Scheme 6.1).[9a] Compound **3** was prepared by esterification of commercially available *m*-nitrophenylacetic acid **1** under Mukaiyama conditions, followed by reduction of the nitro group to hydroxylamine with Zn dust in

$NH_4Cl$ solution and immediate reoxidation with $FeCl_3$ to the nitroso compound. Aniline **6** was obtained from commercially available 3-nitrobenzylamine hydrochloride **4** via a two-step sequence, involving introduction of an Fmoc protecting group followed by reduction of the nitro group. Coupling of **3** and **6** in acetic acid afforded the fully protected azobenzene amino acid **7**, and subsequent cleavage of the *tert*-butyl ester group with TFA gave the turn mimetic amino acid **8**.




**Scheme 6.1.** Synthesis of azobenzene amino acid **8**.

Compound **8** was well suited for standard Fmoc-solid phase peptide synthesis. In the synthesis of the full hairpin, the WKWK sequence was left unaltered other than the two Asn-Gly turn residues, which were both replaced with the single azobenzene amino acid **8** (Figure 6.2).



**9**

**Figure 6.2.** Structure of the WKWK β-hairpin peptide **9** containing a *cis*-azobenzene turn structure.

**ii. Biophysical characterization of a photoswitchable WKWK peptide.** The peptide was first examined to determine the percent composition of *trans*-**9** and *cis*-**9** at thermodynamic equilibrium. A peptide solution was prepared, and as judged by LC/MS at 220 nm, it was determined that approximately 75% of the peptide existed in the trans configuration, whereas 25% adopted the cis configuration. The two isomers were well resolved and easily identified, as the trans isomer is known to be thermodynamically favored and present in higher concentrations. The expected strong absorbance of *trans*-**9** at 320 nm in comparison to *cis*-**9** was also observed, as the peptide appears to have almost exclusively a trans configuration at this wavelength.

Conversion of the hairpin to its photostationary state was then investigated upon exposure to 365 nm light. The length of exposure required for complete conversion to the photostationary state was first evaluated by UV-Vis. After one minute of exposure, a significant decrease in absorbance at 320 nm was observed, corresponding to a decrease in the amount of *trans*-**9** present in solution (Figure 6.3). Similarly, a more subtle

increase in absorbance at 420 nm was observed, indicative of an increase in concentration of *cis*-**9**. To determine if the hairpin was fully converted to the photostationary state, the peptide solution was exposed to 365 nm light for another 5 minutes and again the composition was assessed by UV-Vis. An even further drop in the absorbance at 320 nm was observed, indicating that 1 min of light exposure was not sufficient to reach the photostationary state. After yet a third isomerization of an additional 5 minutes, no further change in the UV spectra was observed, and it was concluded that about 5 minutes of photolysis was required for the peptide to achieve its photostationary state. LC/MS analysis indicated that the hairpin had been converted to a ratio of about 85% *cis*-**9** to 15% *trans*-**9** at 220 nm.

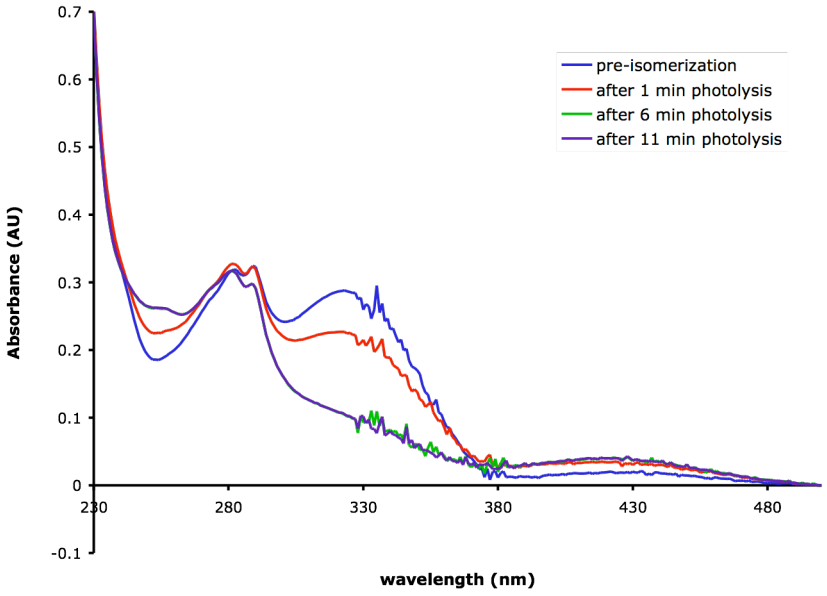



**Figure 6.3.** UV-Vis spectra of hairpin **9** at thermodynamic equilibrium (blue), after photolysis for one minute (red), 6 minutes (green) and 11 minutes (purple).

To confirm the successful isomerization of the hairpin, the peptide was characterized by NMR in both the thermodynamic and photostationary state at room temperature.

Despite overlap with the aromatic tryptophan residues, distinct differences were observed in the aromatic regions of the spectra. The ratio of *cis*-**9**:*trans*-**9** in each spectra, prior to and after photolysis, agree well with the corresponding ratios determined by LC/MS. Furthermore, the *cis*-azobenzene protons were upfield shifted from those of the *trans*-azobenzene moiety, indicating a distinct change in environment of the turn mimetic upon photolysis, and potentially some degree of interaction between the two rings when they are brought in close proximity. The tryptophan α-protons ($H_\alpha$) were also found to be slightly downfield shifted in the photostationary state, which is indicative of a higher degree of β-sheet structure.[13] In conjunction with 2D TOSCY spectra, the *cis*-azobenzene peptide was determined to be approximately 20 – 30% folded, while the trans conformation showed no signs of folding. Both Trp residues in *trans*-**9** show identical chemical shifts, indicating that they are in near identical environments and thus the peptide is likely fully unfolded.

After having characterized the isomerization of the azobenzene chromophore within the WKWK sequence, the structure of the peptide was investigated in both the thermodynamic and photostationary state by circular dichroism. The peptide was first evaluated in the thermodynamic state where the trans isomer is dominant, and a distinct minimum at 200 nm was observed which is characteristic of random coil peptides (Figure 6.4). This indicates that replacement of the Asn-Gly turn of the native WKWK with a *trans*-azobenzene residue effectively disrupts the hairpin structure. A comparable sample isomerized to the photostationary state, where *cis*-**9** was now the dominant species by about 85:15, shows a quite similar spectra. The minimum at about 200 nm indicates that the peptide remains predominantly unfolded when the majority of the peptide takes on a

*cis*-azobenzene β-turn mimic. However, a slight maximum is now observed at 230 nm, which is due to the contribution of Trp-Trp exciton coupling, or stacking of the cross-strand tryptophan residues. This provides evidence that there is a small degree of folding initiated by photolysis. While this may not seem significant, it does not seem extremely dissimilar from the corresponding photoswitchable TrpZip system, which exhibits only a 25% hairpin content of the *cis*-azo isomer at 30 °C as determined by CD.

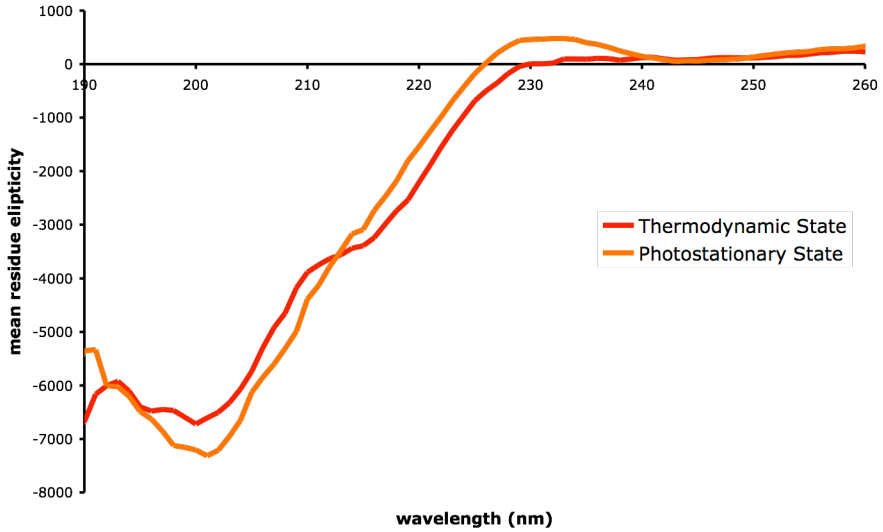



**Figure 6.4.** Circular dichroism spectra for azobenzene peptide **9** at thermodynamic equilibrium (red) and in the photostationary state (orange). Studies were performed in 10 mM sodium phosphate buffer pH 5.5 at 25 °C.

**iii. ATP recognition by a photoswitchable WKWK peptide.** Similar experiments to those described above were performed in the presence of ATP, in both the thermal and photostationary state, to determine the effect of possible ATP recognition on both the peptide's structure and function. While the native WKWK sequence effectively recognizes ATP with micromolar affinity, it naturally exists in a pre-organized state. In comparison, in the cis conformation **9** is only weakly folded, and entirely unfolded in the trans conformation. We sought to investigate what effect, if any, the presence of ATP

had on either the cis:trans ratio of the peptide or the structure of the peptide, and whether binding to ATP could indeed be photocontrolled.

A 3-fold excess of ATP was first added to a low concentration sample of **9** at thermal equilibrium. When directly compared to a corresponding sample without ATP by LC/MS, no change in the ratio of *cis*-**9** to *trans*-**9** was observed, and both contained *trans*-**9** as the dominant isomer in solution (Figure 6.5). Similarly, the ratio of *cis*-**9** to *trans*-**9** in the photostationary state remained unchanged in the presence of ATP. Regardless of whether ATP was added prior to or after the photolysis of **9**, the amount of peptide in the cis conformation in solution remained at about 85%.

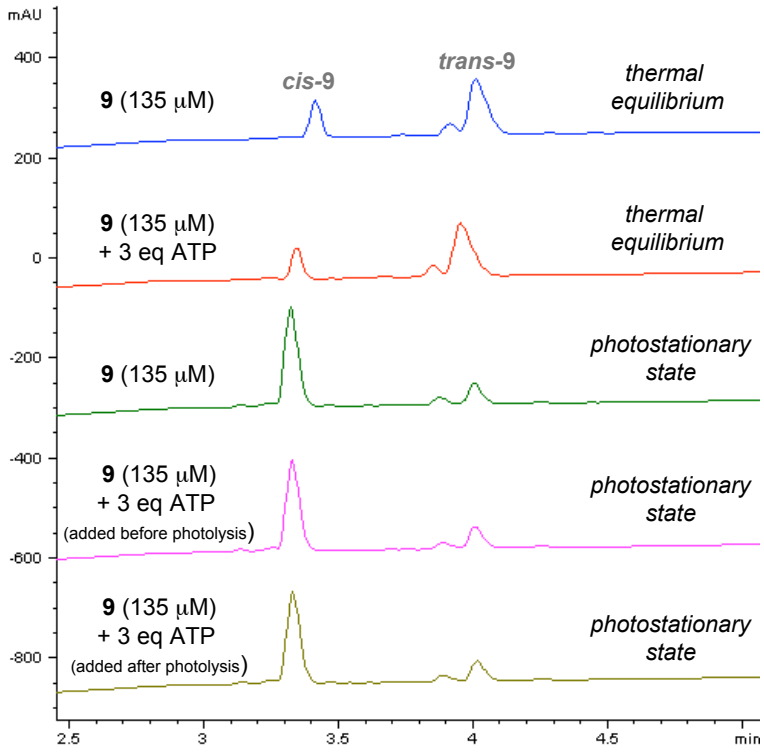



**Figure 6.5.** HPLC traces at 220 nm of azobenzene containing hairpin **9** at thermodynamic equilibrium (blue), at thermodynamic equilibrium with 3 equivalents of ATP (red), in the photostationary state (green), in the photostationary state with 3 equivalents of ATP added prior to photolysis (magenta), in the photostationary state with 3 equivalents of ATP added after photolysis (gold). The peak at 3.9 minutes in each trace corresponds to a small peptide impurity.

To evaluate if the addition of ATP had any affect on the hairpin structure, CD was used to evaluate both the trans and cis peptide conformations in the presence of ATP. Surprisingly, the spectra of both samples containing either predominantly *cis*-**9** or *trans*-**9** were altered significantly by the presence of ATP, and similar changes were observed in each case. When ATP was added to **9** at thermal equilibrium, a distinct new minimum around 215 was observed, which is indicative of β-sheet structure (Figure 6.6). At the same time, the signal around 200 nm corresponding to the random coil conformation is significantly diminished. This indicates that some degree of structural rearrangement upon the addition of ATP is occurring, however without reconfiguration of the azobenzene unit to the cis isomer, which would be expected to facilitate hairpin formation. To confirm that the observed changes were not due simply to the presence of ATP, the CD spectra of ATP at the same concentration was investigated in isolation, showing a much less pronounced signal at about 215 nm and verifying that the new minimum observed is predominately due to a change in the peptide structure.

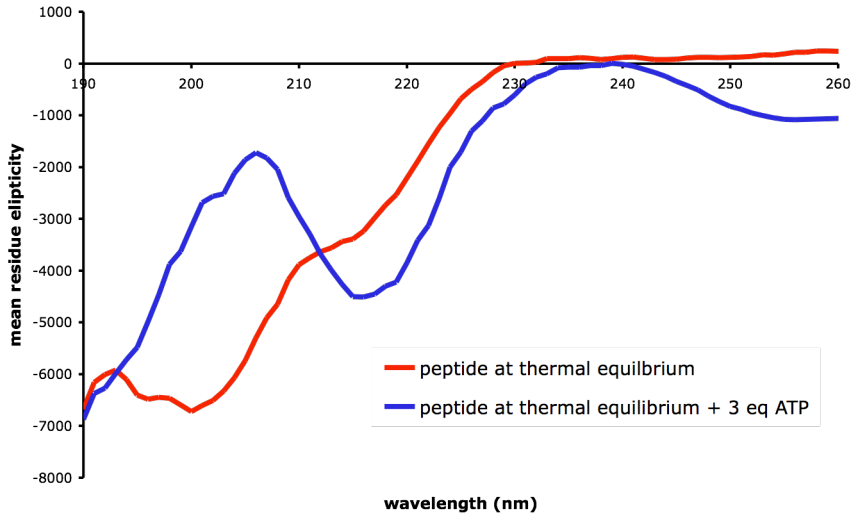



**Figure 6.6.** Circular dichroism spectra for azobenzene peptide **9** at thermodynamic equilibrium (red) and in the presence of ATP (blue). Studies were performed in 10 mM sodium phosphate buffer pH 5.5 at 25 °C.

A similar change in the CD spectra is observed when ATP is added to **9** in the photostationary state. The distinct minimum at 215 is again observed with a concomitant loss in the 200 nm signal, indicating a conversion of the random coil peptide to a β-sheet like structure (Figure 6.7). In addition, a more noticeable maximum signal around 230 is observed, again indicating an interaction between the two cross strand tryptophan residues, which is characteristic of such β-hairpin sequences. A similar CD spectrum is observed for the peptide regardless of whether the ATP is added prior to or after photolysis. While the peptide starts out quite weakly folded, it is appears that the addition of ATP initiates a binding interaction, which induces a greater degree of peptide folding. As the binding of ATP by the native WKWK sequence is largely a result of the diagonal relationship of the two Trp residues on opposite strands which form a binding cleft for aromatic recognition, it is likely that upon addition of ATP in this case, the two peptide strands are brought in closer proximity resulting in an overall more well folded system. This is in contrast to the native WKWK sequence, where the peptide is pre-organized for nucleotide recognition. It is evident that by replacing the well-defined type I' Asn-Gly turn with this more flexible azobenzene unit, a binding induced folding of the hairpin is facilitated.

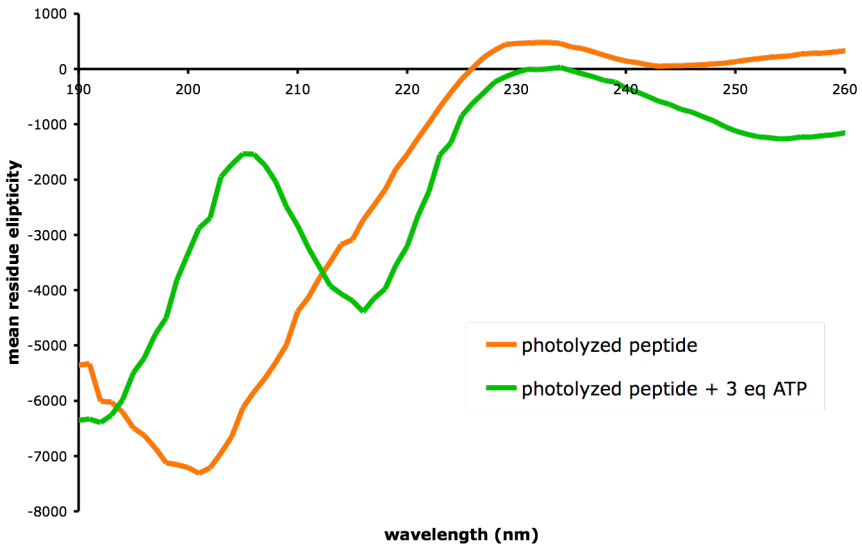



**Figure 6.7.** Circular dichroism spectra for azobenzene peptide **9** in the photostationary state (orange) and in the presence of ATP (green). Studies were performed in 10 mM sodium phosphate buffer pH 5.5 at 25 °C.

In attempt to gain greater insight as to the structural features of the peptide in both the thermal and photostationary state in the presence of ATP, we hoped to perform NMR studies to further examine the system. Upon the addition of excess ATP to concentrations of **9** required for NMR studies, a significant amount of precipitation was observed, preventing further analysis. While this could be corrected for with the addition of a significant amount of a co-solvent such as deuterated acetonitrile, ATP binding is expected to be significantly suppressed in this case, resulting in minimal changes of the peptide spectra in the presence of ATP.

**iv. Peptide mutations.** To evaluate whether the binding induced folding of the designed hairpin was indeed specific for this particular system, various mutations were made to the hairpin sequence to investigate if similar structural changes were observed upon the addition of ATP. First, the azobenzene turn residue was replaced with two glycine residues, which are entirely flexible and have no precedence as a turn sequence.

By removing the azobenzene moiety which is capable of mimicking a β-turn, it was investigated whether folding could be induced by nucleotide recognition in any unfolded system where the sequence of the antiparallel strands was maintained. Investigation of this mutated peptide by CD, both in the absence and presence of ATP, showed identical spectra in both cases (Figure 6.8). This indicated that the azobenzene turn moiety is in fact required to induce a structural change in the peptide upon nucleotide recognition.

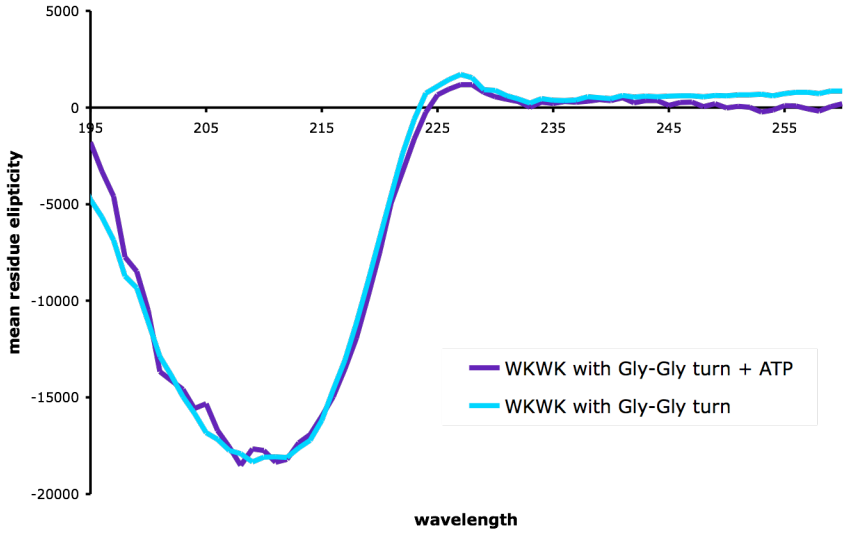



**Figure 6.8.** Circular dichroism spectra for WKWK with a Gly-Gly turn (blue) and in the presence of ATP (purple). Studies were performed in 10 mM sodium phosphate buffer pH 5.5 at 25 °C.

Next, to confirm that ATP is indeed intercalating between the two cross strand Trp residues as in the native sequence, both Trp residues were mutated. The tryptophans were replaced with hydrophobic, aliphatic leucine residues, which are also known for having a reasonably high β-sheet propensity. Evaluation of this control peptide also ensures that the observed changes with hairpin **9** are not the result of non-specific interactions between ATP and the azobenzene group. Initial investigation of this peptide at thermodynamic equilibrium, prior to the addition of ATP, revealed a random coil

conformation, and a spectrum similar to that observed previously for **9** with the expected minimum at 200 nm (Figure 6.9). Upon photolysis, a very similar CD spectra was obtained, indicating that isomerization of the azobenzene unit to the cis confirmation does not result in a well folded peptide. In contrast to *cis*-**9**, no maximum at 230 nm is observed due to the lack of tryptophan residues in this sequence.

Addition of ATP to the LKLK peptide in both the thermal and photostationary state revealed only minor changes in both spectra. In each case the random coil signal is slightly weaker, and only a minor increase in signal is observed around 215 nm. Despite the fact that the tryptophan residues are not present to facilitate ATP binding, both lysine residues remain intact, which are known to interact with the phosphate groups of ATP via favorable electrostatic interactions. As a result, it is likely that ATP is still interacting to some degree with the LKLK control peptide in this fashion, resulting in a minor change in the CD spectra of both the cis and trans peptide. This confirms that the ATP is in fact interacting between the diagonal Trp residues of **9**, as removal of these side chains significantly weakens the binding induced folding of the β-hairpin.

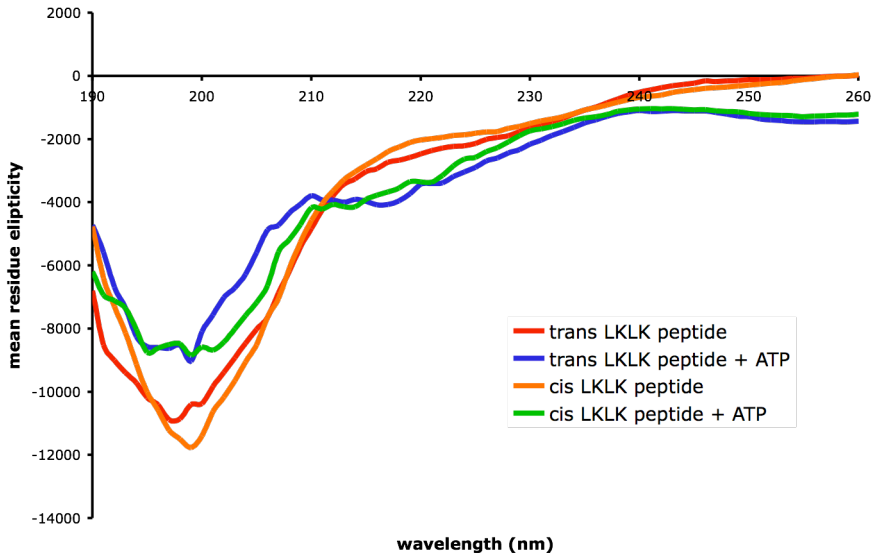



**Figure 6.9.** Circular dichroism spectra for the LKLK azobenzene peptide at thermodynamic equilibrium (red), at thermodynamic equilibrium with ATP (blue), in the photostationary state (orange), and in the photostationary state with ATP (green). Studies were performed in 10 mM sodium phosphate buffer pH 5.5 at 25 °C.

## C. Conclusion

In conclusion, we have designed a β-hairpin peptide whose secondary structure can be induced by its ability to recognize and bind ATP. The hairpin sequence is based on the known nucleotide receptor peptide, WKWK, however the type I' Asn-Gly turn was replaced with a photoswitchable azobenzene moiety. This allows for photochemical control of the peptide structure through photolysis of the azobenzene turn, converting it from completely unfolded in the trans confirmation, to only modestly folded in the cis conformation. Upon addition of ATP, the peptide takes on a β-sheet structure in both the thermodynamic and photostationary state. It is possible, as the azobenzene turn in the trans conformation is not expected to nucleate hairpin formation, that the β-sheet formation in the case of *trans*-**9** is in fact intermolecular, as opposed to intramolecular in the case of *cis*-**9**. Both the azobenzene turn and the cross strand tryptophan residues are

required to facilitate this binding induced folding phenomenon, as mutation of each separately prevents β-sheet formation upon addition of ATP. Gaining a greater understanding of protein folding and being able to control peptide and protein structure with the use of an external trigger is quite valuable, as often the change in a protein's secondary structure can result in vastly different functions, which in some cases proves to be deleterious. We have demonstrated a unique system that enables such structural control based on the fundamental principles of molecular recognition.

**D. Experimental Section**

**i. Peptide synthesis.** All peptide synthesis was performed on an Applied Biosystems Pioneer Peptide Synthesizer using a PEG-PAL-PS resin. Peptides were synthesized on a 0.1 mmol scale by standard Fmoc solid phase synthesis. All amino acids with functionality were protected during synthesis. Coupling reagents were HOBt/HBTU in DMF. For the coupling of arginine, Fmoc-Arg(Boc)2-OH was used in place of the normal Pbf protected residue, as side reactions of unknown nature involving the azobenzene moiety have been reported in the acidolytic removal of the Pbf group from arginine.[14] For the coupling of **8**, only 1.5 equivalents were used and the coupling was performed for 10-12 hours off of the synthesizer. Fmoc deprotections were carried out in 20% piperidine in DMF and, the N-terminus was acylated with a solution of 5% acetic anhydride and 6% 2,6-lutidine in DMF. Cleavage of the peptide from the resin and of the side chain protecting groups was performed by hand with a cocktail of 95% TFA/5% $H_2O$ for 3 hours. TFA was evaporated and the product was precipitated with cold ether. The water-soluble peptides were extracted with water and lyophilized. Peptides were

purified by semipreparative reversed-phase HPLC on a C18 column at a flow rate of 4 mL/min. Peptides were purified with a linear gradient of A and B (A: 95% $H_2O$/5% $CH_3CN$ with 0.1% TFA, B: 95% $CH_3CN$/5% $H_2O$ with 0.1% TFA) and elution was monitored at 214 nm, as well as 320 nm in the case of **9**. Generally only the *trans*-**9** was collected due to its much higher concentration and *cis*-**9** was often found to co-elute with other peptide impurities. Once purified, peptides were lyophilized to powder to give a bright orange solid in the case of **9**, and characterized by +ESI-MS.

**ii. Peptide photoisomerization.** Azobenzene containing peptides were irradiated at 365 nm for five minutes in the dark with a Spectroline long wave UV pencil lamp (1,000 $\mu W/cm^2$ of 365 nm radiation at 1"). Once irradiated, solutions were continually stored in the dark at room temperature. Analysis and characterization of the photolyzed peptides was generally carried out immediately after photolysis. A maximum of *cis*- to *trans*-azo ratios of about 85:15 are obtained at the photostationary state as determined by LC/MS.

**iii. UV-Vis measurements.** A sample of hairpin **9** was prepared at a concentration of 45 μM as determined by mass in NaOAc buffer, pH 3.7. UV scans were taken over 500 to 200 nm, both prior to and after photolysis.

**iv. Analytical LC/MS.** LC-MS was carried out on an Agilent Rapid Resolution LC-MSD system equipped with an online degasser, binary bump, autosampler, heated column compartment, and diode array detector. All separations were performed using $H_2O$-acetonitrile gradients with 0.2% formic acid and a Zorbax Eclipse XDB-C18 column (4.6 × 50 mm, 1.8 micron). The MS was performed using a single quad mass spectrometer. Mass spectra (ESI+) were acquired in ultrascan mode by using a drying temperature of 350°C, a nebulizer pressure of 45 psi, a drying gas flow of 10 L/min, and

a capillary voltage of 3000 V. Injection volumes were varied from 4 µl to 15 µl depending on the sample concentration, in order to achieve adequate MS and UV detection. The chromatography of was carried out at 45°C at a flow rate of 1.5 mL/min (0-20% B, 0-2 minutes / 20-31% B, 2-6 minutes). Multiple wavelengths were monitored for analysis (220 nm, 280 nm, 320 nm) and the peak areas were integrated at 220 nm. The cis and trans isomers were assigned based on a strong absorbance of trans azobenzene at 320 nm, along with the shift in the distribution of isomers upon isomerization.

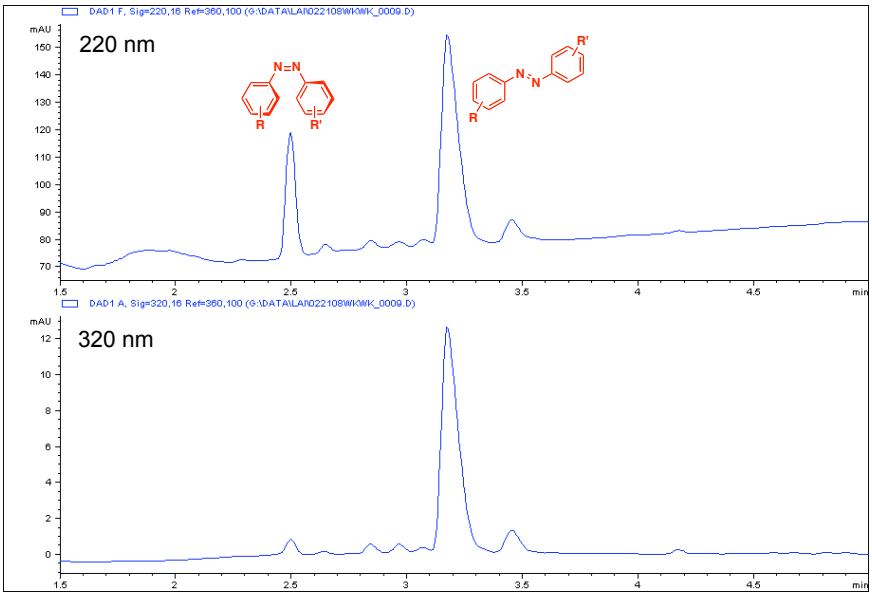



**Figure 6.10.** LC/MS trace of azobenzene containing hairpin **9** at thermodynamic equilibrium at 220 nm (top) and 320 nm (bottom). The cis:trans ratio of the peptide was determined to be 25:75 at 220 nm.

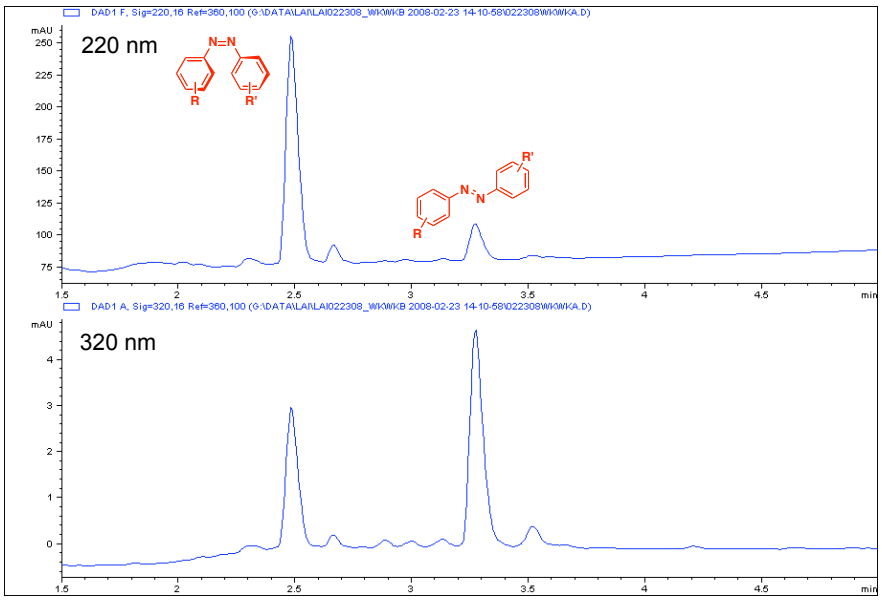



**Figure 6.11.** LC/MS trace of azobenzene containing hairpin **9** in the photostationary state at 220 nm (top) and 320 nm (bottom). The cis:trans ratio of the peptide was determined to be 85:15 at 220 nm.

**v. NMR spectroscopy.** Peptide NMR samples were prepared in concentrations of 0.7 – 1.0 mM and analyzed on a Varian Inova 600-MHz instrument at 25 °C. Samples were dissolved in one third $CD_3CN$ and two-thirds $D_2O$ buffered to pD 4.0 with 50 mM NaOAc and 240 mM AcOD with 0.5 mM DSS. 1D NMR spectra were collected using 32K data points and 32 scans using a 1-3 second presaturation. All 2D NMR experiments used pulse sequences from the Chempack software. 2D TOSCY experiments were taken with 16 scans in the first dimension and 128 scans in the second dimension. All spectra were analyzed using standard window functions (sinbell and Gaussian with shifting). Presaturation was used to suppress the water resonance.

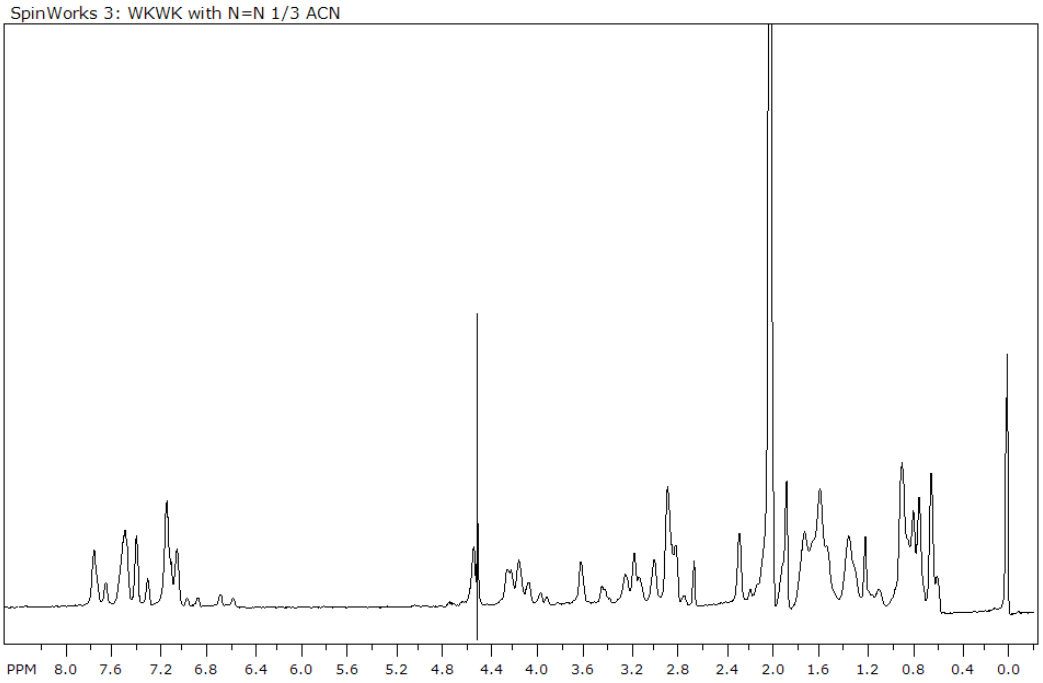



**Figure 6.12.** $^{1}$H NMR of **9** at thermodynamic equilibrium (25% cis, 75% trans) in one-third $CD_3CN$ at 25 °C.

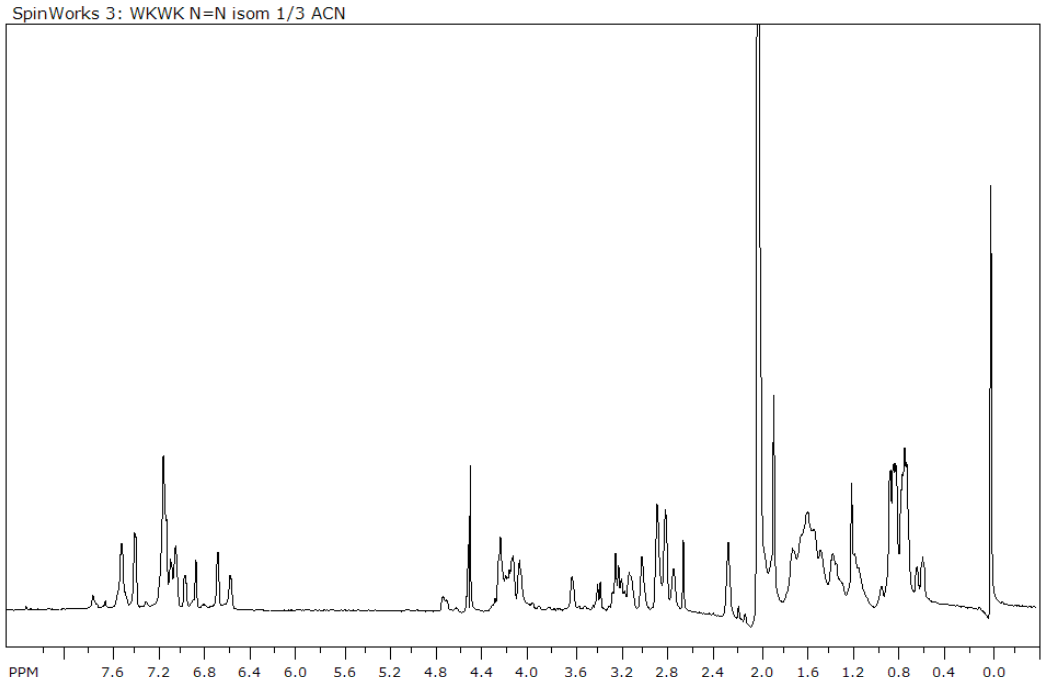



**Figure 6.13.** $^{1}$H NMR of **9** in the photostationary states (85% cis, 15% trans) in one-third $CD_3CN$ at 25 °C.

**vi. Circular dichroism.** CD spectra in the range of 185 to 500 nm were recorded on a Pistar-180 spectropolarimeter equipped with a thermostated cell holder. In the case of the control Gly-Gly turn peptide, the spectra were collected on an Aviv Model 62DS CD spectrometer. In each case, spectra were obtained at 25 °C with a 1.0 nm step, a bandwidth of 2 nm, and 2.5 seconds per point in a 0.1 cm quartz cell. Spectra were generally taken at peptide concentrations of either 45 μM or 135 μM as determined by mass in phosphate buffer (pH 5.5). The error associated with determining peptide concentrations by mass is quite large, resulting in CD signals of varying intensities. When appropriate, three equivalents of ATP were added as determined by UV absorption at 259 nm ($\varepsilon$ = 15,400 $M^{-1}cm^{-1}$). A buffer subtraction was performed in each case, and the results are reported in terms of molar ellipticity per residue:

$$[\Theta]_R = (\Theta_{obs} / (10 \times l \times c)) / r$$

where $\Theta_{obs}$ is the observed signal, $l$ is the pathlength (0.1 cm), c is the peptide concentration, and r is the number of residues (11 was used for the azobenzene peptides). The CD spectra of the *cis*-azo isomers were recorded immediately after irradiation at 365 nm until the photostationary state was reached.

[1] Gellman, S. H. *Curr. Opin. Struct. Biol.* **1998**, *2*, 717-725.

[2] Griffiths-Jones, S. R.; Maynard, A. J.; Sharman, G. J.; Searle, M. S. *Chem. Commun.* **1998**, 789-790.

[3] (a) Butterfield, S. M.; Waters, M. L. *J. Am. Chem. Soc.* **2003**, *125*, 9580-9581. (b) Butterfield, S. M.; Sweeney, M. M.; Waters, M. L. *J. Org. Chem.* **2005**, *70*, 1105-1114.

[4] Stanger, H. E.; Gellman, S. H. *J. Am. Chem. Soc.* **1998**, *120*, 4236-4237.

[5] Renner, C.; Moroder, L. *ChemBioChem*, **2006**, *7*, 868-878.

[6] For example: (a) Pieroni, O.; Houben, J. L.; Fissi, A.; Costantino, P. *J. Am. Chem. Soc.* **1980**, *102*, 5913-5915. (b) Kumita, J. R.; Smart, O. S.; Woolley, G. A. *Proc. Natl, Acad, Sci. USA* **2000**, *97*, 3803-3808.

[7] For example: Renner, C.; Behrendt, R.; Sporlein, S.; Wachtveitl, J.; Moroder, L. *Biopolymers* **2000**, *54*, 489-500.

[8] Ulysse, L.; Cubillos, J.; Chmielewski, J. *J. Am. Chem. Soc.* **1995**, *117*, 8466-8467.

[9] (a) Aemissegger, A.; Krautler, V.; van Gunsteren, W. F.; Hilvert, D. *J. Am. Chem. Soc.* **2005**, *127*, 2929-2936. (b) Dong, S.-L.; Loweneck, M.; Schrader, T. E.; Schreier, W. J.; Zinth, W.; Moroder, L.; Renner, C. *Chem Eur. J.* **2006**, *12*, 1114-1120.

[10] (a) Russell, S. J.; Cochran, A. G. *J. Am. Chem. Soc.* **2000**, *122*, 12600-12601. (b) Cochran, A. G.; Tong, R. T.; Starovasnik, M. A.; Park, E. J.; McDowell, R. S.; Theaker, J. E.; Skelton, N. J. *J. Am. Chem. Soc.* **2001**, *123*, 625-632. (c) Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 5578-5583.

[11] Pieroni, O.; Fissi, A.; Angelini, N.; Lenci, F. *Acc. Chem. Res.* **2001**, *34*, 9-17.

[12] Krautler, V.; Aemissegger, A.; Hunenberger, P. H.; Hilvert, D.; Hansson, T.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **2005**, *127*, 4935-4942.

[13] Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1992**. *31*, 1647-1651.

[14] Renner, C.; Kusebauch, U.; Loweneck, M.; Milbradt, A. G.; Moroder, L. *J. Peptide Res.* **2005**, *65*, 4-14.

## References


Abell, A. D.; Jones, M. A.; Neffe, A. T.; Aitken, S. G.; Cain, T. P.; Payne, R. J.; McNabb, S. B.; Coxon, J. M.; Stuart, B. G.; Pearson, D.; Lee, H. Y-Y.; Morton, J. D. *J. Med. Chem.* **2007**, *50*, 2916-2920.

Aemissegger, A.; Krautler, V.; van Gunsteren, W. F.; Hilvert, D. *J. Am. Chem. Soc.* **2005**, *127*, 2929-2936.

Albrecht, M.; Blau, O.; Frohlich, R. *Chem. Eur. J.* **1999**, *5*, 48-56.

Arents, G.; Burlingame, R. W.; Wang, B. C.; Love, W. E.; Moudrianakis, E. N. *Proc.* Natl. Acad. Sci. **1991**, 88, 10148-10152.

Ball, L. J.; Kuhne, R.; Schneider-Mergener, J.; Oschkinat, H. *Angew. Chem. Int. Ed.* **2005**, *44*, 2852-2869.

Bang, D.; Pentelute, B. L.; Gates, Z. P.; Kent, S. B. *Org. Lett.* **2006**, *8*, 1049-1052.

Bannister, A. J.; Kouzarides, T. *Nature* **2005**, *436*, 1103-1106.

Baolu, S.; Stevenson, R.; Campopiano, D. J.; Greaney, M. F. *J. Am. Chem. Soc.* **2006**, *128*, 8459-8467.

Baxter, P. N. W.; Khoury, R. G.; Lehn, J.-M.; Baum, G.; Fenske, D. *Chem. Eur. J.* **2000**, *6*, 4140-4148.

Bedford, M. T.; Richard, S. *Molecular Cell* **2005**, *18*, 263-272.

Berger, S. L. *Nature* **2007**, *447*, 407-412.

Berkovich-Berger, D.; Lemcoff, N. G. *Chem. Commun.* **2008**, 1686-1688.

Besenius, P.; Cormack, P. A. G.; Liu, J.; Otto, S.; Sanders, J. K. M.; Sherrington, D. C. *Chem. Eur. J.* **2008**, *14*, 9006-9019.

Beshara, C. S.; Jones, C. E.; Daze, K. D.; Lilgert, B. J.; Hof, F. *ChemBioChem* **2010**, *11*, 63-66.

Bhaumik, S. R.; Smith, E.; Shilatifard, A. *Nature Struct. Molec. Biol.* **2007**, *14*,1008-1016.

Blow, N. *Nature* **2007**, *447*, 741–744.

Bornaghi, L. F.; Wilkinson, B. L.; Kiefel, M. J.; Poulsen, S. A. *Tettrahedron Lett.* **2004**, *45*, 9281-9284.

Bottomley, M. J. *Embo Reports* **2004**, *5*, 464-469.

Boul, P. J.; Reutenauer, P.; Lehn, J.-M. *Org. Lett.* **2005**, *7*, 15-18.

Boulegue, C.; Loweneck, M.; Renner, C.; Moroder, L. *ChemBioChem* **2007**, *8*, 591-594.

Boyne, M. T.; Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L. *J. Proteome Res.* **2006**, *5*, 248-253.

Brady, P. A.; Sanders, J. K. M. *J. Chem. Soc. Perkin I* **1997**, 3237-3253.

Brik, A.; Yang, Y. Y.; Ficht, S.; Wong, C. H. *J. Am. Chem. Soc.* **2006**, *128*, 5626-5627.

Bugaut, A.; Toulmé, J.-J.; Rayner, B. *Org. Biomol. Chem.* **2006**, *4*, 4082-4088.

Bulos, F.; Roberts, S. L.; Furlan, R. L. E.; Sanders, J. K. M. *Chem. Commun.* **2007**, 3092-3093.

Burley, S. K.; Petska, G. A. *FEBS Letters* **1986**, *203*, 139-143.

Butterfield, S. M.; Sweeney, M. M.; Waters, M. L. *J. Org. Chem.* **2005**, *70*, 1105-1114.

Butterfield, S. M.; Waters, M. L. *J. Am. Chem. Soc.* **2003**, *125*, 9580-9581.

Cacciapaglia, R.; Di Stefano, S.; Mandolini, L. *J. Am. Chem. Soc.* **2005**, *127*, 13666-13671.

Choudhary, S.; Morrow, J. R. *Angew. Chem. Int. Ed.* **2002**, *41*, 4096-4098.

Christinat, N.; Scopelliti, R.; Severin, K. *Angew. Chem. Int. Ed.* **2008**, *47*, 1848-1852.

Christinat, N.; Scopelliti, R.; Severin, K. *J. Org. Chem.* **2007**, *72*, 2192-2200.

Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 5578-5583.

Cochran, A. G.; Tong, R. T.; Starovasnik, M. A.; Park, E. J.; McDowell, R. S.; Theaker, J. E.; Skelton, N. J. *J. Am. Chem. Soc.* **2001**, *123*, 625-632.

Corbett, P. T.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J-L.; Sanders, J. K. M.; Otto, S. *Chem. Rev.* **2006**, *106*, 3652-3711.

Corbett, P. T.; Otto, S.; Sanders, J. K. M. *Chem. Eur. J.* **2004**, *10*, 3139.

Corbett, P. T.; Otto, S.; Sanders, J. K. M. *Org. Lett.* **2004**, *6*, 1825-1827.

Corbett, P. T.; Sanders, J. K. M.; Otto, S. *Angew. Chem. Int. Ed.* **2007**, *46*, 8858-8861.

Corbett, P. T.; Sanders, J. K. M.; Otto, S. *Chem. Eur. J.* **2008**, *14*, 2153-2166.

Corbett, P. T.; Sanders, J. K. M.; Otto, S. *J. Am. Chem. Soc.* **2005**, *127*, 9390-9392.

Corbett, P. T.; Tong, L. H.; Sanders, J. K. M.; Otto, S. *J. Am. Chem. Soc.* **2005**, *127*, 8902-8903.

Cosgrove, M. S.; Boeke, J. D.; Wolberger, C. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1037-1043.

Cote, J.; Richard, S. *J. Biol. Chem.* **2005**, *280*, 28476-28483.

Cousins, G. R. L.; Poulsen, S. A.; Sanders, J. K. M. *Chem. Commun.* **1999,** 1575-1576.

Cowling, V. H. *Biochem J.* **2010**, *425*, 295-302.

Craik, D. J. *Science* **2006**, *311*, 1563-1564.

Crego Calama, M.; Hulst, R.; Fokkens, R.; Nibbering, N. M. M.; Timmerman, P.; Reinhoudt, D. N. *Chem. Commun.* **1998**, 1021-1022.

Crespo, L.; Sanclimens, G.; Montaner, B.; Perez-Tomas, R.; Royo, M.; Pons, M.; Albericio, Giralt, E. *J. Am. Chem. Soc.* **2002**, *124*, 8876-8883.

Daniel, J. A.; Pray-Grant, M. G.; Grant, P. A. *Cell Cycle* **2005**, *4*, 919-926.

Darzynkiewicz, E.; Stepinski, J.; Tahara, S. M.; Stolarski, R.; Ekiel, I.; Haber, D.; Neuvonen, K.; Lehikonen, P.; Labadi, I.; Lonnberg, H. *Nucleosides Nucleotides* **1990**, *9*, 599-618.

De Benedetti, A.; Graff, J. R. *Oncogene* **2004**, *23*, 3189.

Dong, S-L.; Loweneck, M.; Schrader, T. E.; Schreier, W. J.; Zinth, W.; Moroder, L.; Renner, C. *Chem. Eur. J.* **2006**, *12*, 1114-1120.

Edelhoch, H. *Biochemistry*, **1967**, *6*, 1948-1954.

Eldred, S. E.; Stone, D. A.; Gellman, S. H.; Stahl, S. S. *J. Am. Chem. Soc.* **2003**, *125*, 3422-3423.

Eliseev, A. V.; Nelen, M. I. *J. Am. Chem. Soc.* **1997**, *119*, 1147-1148.

Erlanson, D. A.; Hansen, S. K. *Curr. Opin. Chem. Biol.* **2004**, *8*, 399.

Fernandes, C.; Correia, J. D. G.; Gano, L.; Santos, I.; Seifert, S.; Syhre, R.; Bergmann, R.; Spies, H. *Bioconjugate Chem.* **2005**, *16*, 660-668.

Field, L.; Engelhardt, P. R. *J. Org. Chem.* **1970**, *35*, 3647-3655.

Fischle, W.; Tseng, B. S.; Dormann, H. L.; Ueberheide, B. M.; Garcia, B. A.; Shabanowitz, J.; Hunt, D. F.; Funabiki, H.; Allis, C. D. *Nature* **2005**, *438*, 1116-1122.

Flocco, M. M.; Mowbray, S. L. *J. Mol. Biol.* **1994**, *235*, 709-717.

Freitas, M. A.; Sklenar, A. R.; Parthun, M. R. *J. Cell Biochem.* **2004**, *92*, 691-700.

Fuchs, B.: Nelson, A.; Star, A.; Stoddart, J. F.; Viidal, S. *Angew. Chem. Int. Ed.* **2003**, *42*, 4220-4224.

Furlan, R. L. E.; Ng, Y.-F.; Otto, S.; Sanders, J. K. M. *J. Am. Chem. Soc.* **2001**, *123*, 8876.

Furlan, R. L. E.; Yiu-Fai, N.; Cousins, G. R. L.; Redman, J. E.; Sanders, J. K. M. *Tetrahedron* **2002**, *58*, 771-778.

Gallivan, J. P.; Dougherty, D. A. *Proc. Nat. Acad. Sci.* **1999**, *96*, 9459-9464.

Gellman, S. H. *Curr. Opin. Struct. Biol.* **1998**, *2*, 717-725.

Gelman, M. A.; Richter, S.; Cao, H.; Umezawa, N.; Gellman, S. H.; Rana, T. M. *Org. Lett*. **2003**, *5*, 3563-3565.

Goral, V.; Nelen, M. I.; Eliseev, A. V.; Lehn, J-M. *Proc. Nat. Acad. Sci.* **2001**, *98*, 1347-1352.

Gorostiza, P.; Volgraf, M.; Numano, R.; Szobota, S.; Trauner, D.; Isacoff, E. Y. *Proc. Nat. Acad. Sci.* **2007**, *104*, 10865-10870.

Griffiths, J. *Chem. Soc. Rev.* **1972**, *1*, 481-493.

Griffiths-Jones, S. R.; Maynard, A. J.; Sharman, G. J.; Searle, M. S. *Chem. Commun.* **1998**, 789-790.

Grote, Z.; Scopelliti, R.; Severin, K.. *Angew. Chem. Int. Ed.* **2003**, *42*, 3821-3825.

Guccione, E.; Bassi, C.; Casadio, F.; Martinato, F.; Cesaroni, M.; Schuchlautz, H.; Luscher, B.; Amati, B. *Nature* **2007**, *449*, 933-937.

Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. *Proc. Natl. Acad. Sci.* **1999**, *96*, 10068-10073.

Hamilton, D. G.; Feeder, N.; Teat, S. J.; Sanders, J. K. M. *New J. Chem.* **1998**, 1019-1021.

Harvey, J. H.; Trauner, D. *ChemBioChem* **2008**, *9*, 191-193.

Hendler, S.; Furer, E.; Srinivasan, P. R. *Biochemistry* **1970**, *9*, 4141-4153.

Hirota, T.; Lipp, J. J.; Toh, B. H.; Peters, J. M. *Nature* **2005**, *438*, 1176-1180.

Hoerter, J. M.; Otte, K. M.; Gellman, S. H.; Cui, Q.; Stahl, S. S. *J. Am. Chem. Soc.* **2008**, *130*, 647-654.

Hof, F.; Nuckolls, C.; Rebek, J. Jr. *J. Am. Chem. Soc.* **2000**, *122*, 4251-4252.

Horswill, A. R.; Benkovic, S. J. *Cell Cycle* **2005**, *4*, 552-555.

Horton, D. A.; Bourne, G. T.; Smythe, M. L. *J. Comput. Aided Mol. Des.* **2002**, *16*, 415-430.

Huang, J.; Berger, S. L. *Curr. Opin. Genet. Dev.* **2008**, *18*, 152-158.

Huc, I.; Lehn, J. M. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 2106-2110.

Hughes, R. M.; Benshoff, M. L.; Waters, M. L. *Chem. Eur. J.* **2007**, *13*, 5753-5764.

Hughes, R. M.; Waters, M. L. *J. Am. Chem. Soc.* **2005**, *127*, 6518-6519.

Hughes, R. M.; Waters, M. L. *J. Am. Chem. Soc.* **2006**, *128*, 12735-12742.

Hughes, R. M.; Wiggins, K. R.; Khorasanizadeh, S.; Waters, M. L. *Proc. Nat. Acad. Sci. USA* **2007**, *104*, 11184-11188.

Iftime, G.; Labarthet, F. L.; Natansohn, A.; Rochon, P. *J. Am. Chem. Soc.* **2000**, *122*, 12646-12650.

Jacobs, S. A.; Khorasanizadeh, S. *Science* **2002**, *295*, 2080-2083.

Jacobsen, H.; Stockmayer, W. H. *J. Chem. Phys.* **1950**, *18*, 1600-1607.

Jensen, O. N. *Nature Rev. Mol. Cell Biol.* **2006**, *7*, 391-403.

Jenuwein, T.; Allis, C. D. *Science* **2001**, *293*, 1074-1080

Jurt, S.; Aemissegger, A.; Guntert, P.; Zerbe, O.; Hilvert, D. *Angew. Chem. Int. Ed.* **2006**, *45*, 6297-6300.

Kamiichi, K.; Doi, M.; Nabae, M.; Ishida, T.; Inoue, M. *J. Chem. Soc. Perkin Trans. II* **1987**, 1739-1745.

Karan, C.; Miller, B. L. *Drug Discov. Today* **2000**, *5*, 67-75.

Kazanecki, C. C.; Kowalski, A. J.; Ding, T.; Rittling, S. R.; Denhardt, D. T. *J. Cell. Biochem.* **2007**, *102*, 925–935.

Kearney, P. C.; Mizoue, L. S.; Kumpf, R. A.; Forman, J. E.; McCurdy, A.; Dougherty, D. A. *J. Am. Chem. Soc.* **1993**, *115*, 9907-9919.

Khorasanizedeh, S. *Cell* **2004**, *116*, 259-272.

Kiehna, S. E.; Waters, M. L. *Protein Sci.* **2003**, *12*, 2657-2667.

Kirmizis, A.; Santos-Rosa, H.; Penkett, C. J.; Singer, M. A.; Verrmeulen, M.; Mann, M.; Bahler, J. Green, R. D.; Kouzarides, T. *Nature* **2007**, *449*, 928-932.

Kouzarides, T. *Cell* **2007**, *128*, 693-705.

Krautler, V.; Aemissegger, A.; Hunenberger, P. H.; Hilvert, D.; Hansson, T.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **2005**, *127*, 4935-4942.

Kretsinger, J. K.; Schneider, J. P. *J. Am. Chem. Soc.* **2003**, *125*, 7907-7913.

Kumar, G. S.; Neckers, D. C. *Chem. Rev.* **1989**, *89*, 1915-1925.

Kumita, J. R.; Smart, O. S.; Woolley, G. A. *Proc. Natl, Acad, Sci. USA* **2000**, *97*, 3803-3808.

Ladame, S. *Org. Biomol. Chem.* **2008**, *6*, 219-226.

Lam, R. T. S.; Belenguer, A.; Roberts, S. L.; Naumann, C.; Jarrosson, T.; Otto, S.; Sanders, J. K. M. *Science* **2005**, *308*, 667-669.

Larsson, R.; Pei, Z.; Ramström, O. *Angew. Chem. Int. Ed.* **2004**, *43*, 3716-3718.

Larsson, R.; Ramstrom, O. *Eur. J. Org. Chem.* **2006,** *2006*, 285-291.

Leclaire, J.; Vial, L.; Otto, S.; Sanders, J. K. M. *Chem. Commun.* **2005**, 1959-1961.

Lehn, J-M. *Chem. Eur. J.* **1999**, *5*, 2455-2463.

Lehn, J-M. *Prog. Polym. Sci.* **2005**, *30*, 814-831.

Li, H.; Fischle, W.; Wang, W.; Duncan, E. M.; Liang, L.; Murakami-Ishibe, S.; Allis, C. D.; Patel, D. J. *Mol. Cell* **2007**, *28*, 667-691.

Liénard, B. M. R.; Selevsek, N.; Oldham, N. J.; Schofield, C. J. *ChemMedChem* **2007**, *2*, 175-179.

Lindahl, T. *Nature*, **1981**, *290*, 363-364.

Liu, J.; West, K. R.; Bondy, C. R.; Sanders, J. K. M. *Org. Biomol. Chem.* **2007**, *5*, 778-786.

Ludlow, R. F.; Liu, J.; Li, H.; Roberts, S. L.; Sanders, J. K. M.; Otto, S. *Angew. Chem. Int. Ed.* **2007**, *46*, 5762-5764.

Ludlow, R. F.; Otto, S. *Chem. Soc. Rev.* **2008**, *37*, 101-108.

Ludlow, R. F.; Otto, S. *J. Am. Chem. Soc.* **2008**, *130*, 12218-12219.

Li, H. T.; Ilin, S.; Wang, W. K.; Duncan, E. M.; Wysocka, J.; Allis, C. D.; Patel, D. J. *Nature* **2006**, *442*, 91-95.

Lee, D. Y.; Teyessier, C.; Strahl, B. D.; Stallcup, M. R. *Endocrine Reviews* **2005**, *26*, 147-170.

Ma, J. C.; Dougherty, D. A. *Chem. Rev.* **1997**, *97*, 1303-1324.

Mal, P.; Schultz, D.; Beyeh, K.; Rissanen, K.; Nitschke, J. R. *Angew. Chem. Int. Ed.* **2008**, *47*, 8297-8301.

Martin, C.; Zhang, Y. *Nature Rev. Molec. Cell Biol.* **2005**, *6*,838-849.

Matsumoto, M.; Nicholas, K. M. *J. Org. Chem.* **2007**, *72*, 9308-9313.

McBride, A. E.; Silver, P. A. *Cell* **2001**, *106*, 5-8.

McNaughton, B. R.; Miller, B. L. *Org. Lett.* **2006**, *8*, 1803-1806.

Meinke, P. T.; Liberator, P. *Curr. Med. Chem.* **2001**, *8*, 211-235.

Miller, B. L.; Klekota, B. *Tetrahedron* **1999**, *55*, 11687-11697.

Minucci, S.; Pelicci, P. G. *Nat. Rev. Cancer* **2006**, *6*, 38-51.

Mitchell, J. B. O.; Nandi, C. L.; McDonald, I. K.; Thornton, J. M.; Price, S. L. *J. Mol. Biol.* **1994**, *239*, 315-331.

Montero, A.; Beierle, J. M.; Olsen, C. A.; Ghadiri, M. R. *J. Am. Chem. Soc.* **2009**, *131*, 3033–3041.

Moodie, S. L.; Mitchell, J. B. O.; Thorton, J. M. *J. Mol. Biol.* **1996**, *263*, 486-500.

Natansohn, A.; Rochon, P. *Chem. Rev.* **2002**, *102*, 4139-4175.

Natarajan, A.; Moerke, N.; Fan, Y.-H.; Chen, H.; Christ, W. J.; Wagner, G.; Halperin, J. A. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2657-2660.

Nazarpack-Kandlousy N.; Nelen, M. I.; Goral, V.; Eliseev, A. V. *J. Org. Chem.* **2002**, *67*, 59-65.

Nazarpack-Kandlousy, N.; Zweigenbaum, J.; Henion, J.; Eliseev, A. V. *J. Comb. Chem.* **1999**, *1*, 199-206.

Ngola, S. M.; Kearney, P. C.; Mecozzi, S.; Russell, K.; Dougherty, D. A. *J. Am. Chem. Soc.* **1999**, *121*, 1192-1201.

Nguyen, R.; Huc, I. *Chem. Commun*. **2003**, 942-943.

Niedzwiecka, A.; Marcotrigiano, J.; Stepinski, J.; Jankowska-Anyszka, M.; Wyslouch-Cieszynska, A.; Dadlez, M.; Gingras, A.-C.; Mak, P.; Darzynkiewicz, E.; Sonenberg, N.; Burley, S. K.; Stolarski, R. *J. Mol. Biol.* **2002**, *319*, 615-635.

Nielsen, A. L.; Oulad-Abdelghani-M.; Ortiz, J. A.; Remboutsika, E.; Chambon, P.; Losson, R. *Mol. Cell* **2001**, *7*, 729-739.

Nielsen, P. R.; Nietlispach, D.; Mott, H. R.; Callaghan, J.; Bannister, A.; Kouzarides, T. ; Murzin, A. G.; Murzina, N. V.; Laue, E. D. *Nature* **2002**, *416*, 103-107.

Nitschke, J. R. *Acc. Chem. Res.*, **2007**, *40*, 103–112.

Nitschke, J. R.; Lehn, J.-M. *Proc. Natl. Acad. Sci.* **2003**, *100*, 11970-11974.

Olsen, C. A.; Ghadiri, M. R. *J. Med. Chem*, **2009**, *52*, 7836–7846.

Otto, S. *Curr. Opin. Drug Disc. Devel.* **2003**, *6*, 509-520.

Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *Drug Discov. Today* **2002**, *7*, 117-125.

Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *J. Am. Chem. Soc.* **2000**, *122*, 12063-12064.

Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *Science* **2002**, *297*, 590-593.

Patel, K.; Borchardt, R. T. *Pharm. Res.* **1990**, *7*, 703-711.

Pérez-Fernández, R.; Pittelkow, M.; Belenguer, A. M.; Sanders, J. K. M. *Chem. Commun.* **2008**, 1738-1740.

Pesavento, J. J.; Kim, Y. B.; Taylor, G. K.; Kelleher, N. L. *J. Am. Chem. Soc.* **2004**, *126*, 3386-3387.

Petti, M. A.; Shepodd, T. J.; Barrans, R. E.; Dougherty, D. A. *J. Am. Chem. Soc.* **1988**, *110*, 6825-6840.

Pieroni, O.; Fissi, A.; Angelini, N.; Lenci, F. *Acc. Chem. Res.* **2001**, *34*, 9-17.

Pieroni, O.; Houben, J. L.; Fissi, A.; Costantino, P. *J. Am. Chem. Soc.* **1980**, *102*, 5913-5915.

Polyakov, V. A.; Nelen, M. I.; Nazarpack-Kandlousy, N.; Ryabov, A. D.; Eliseev, A. V. *J. Phys. Org. Chem.* **1999**, *12*, 357-363.

Ramierz-Alvarado, M.; Kortemme, T.; Blanco, F. J.; Serrano, L. *Bioorg. Med. Chem.* **1999**, *7*, 93-103.

Ramström, O.; Lehn, J.-M. *ChemBioChem* **2000**, *1*, 41-48.

Renner, C.; Behrendt, R.; Sporlein, S.; Wachtveitl, J.; Moroder, L. *Biopolymers* **2000**, *54*, 489-500.

Renner, C.; Kusebauch, U.; Loweneck, M.; Milbradt, A. G.; Moroder, L. *J. Peptide Res.* **2005**, *65*, 4-14.

Renner, C.; Moroder, L. *ChemBioChem* **2006**, *7*, 868-878.

Richter, J. D.; Sonenberg, N. *Nature*, **2005**, *433*, 477-480.

Roberts, S. L.; Furlan, R. L. E.; Cousins, G. R. L.; Sanders, J. K. M. *Chem. Commun.* **2002**, 938-939.

Roberts, S. L.; Furlan, R. L. E.; Otto, S.; Sanders, J. K. M. *Org. Biomol. Chem.* **2003**, *1*, 1625-1633.

Rodriguez-Docampo, Z.; Otto, S. *Chem. Commun.* **2008**, 5301-5305.

Rowan, S. J.; Hamilton, D. G.; Brady, P. A.; Sanders, J. K. M. *J. Am. Chem. Soc.*, **1997**, *119*, 2578–2579.

Rowan, S. J.; Reynolds, D. J.; Sanders, J. K. M. *J. Org. Chem.* **1999**, *64*, 5804-5814.

Russell, S. J.; Cochran, A. G. *J. Am. Chem. Soc.* **2000**, *122*, 12600-12601.

Ruszczynska, K.; Kamienska-Trela, K.; Wojcik, J.; Stepinski, J.; Darzynkiewicz, E.; Stolarski, R. *Biophysical Journal*, **2003**, *85*, 1450-1456.

Ruthernburg, A. J.; Allis, C. D.; Wysocka, J. *Cell* **2007**, *25*, 15-30.

Salonen, L. M.; Bucher, C.; Banner, D. W.; Haap, W.; Mary, J.-L.; Benz, J.; Kuster, O.; Seiler, P. I.; Schweizer, W. B.; Diederich, F. *Angew. Chem. Int. Ed.* **2009**, *48*, 811-814.

Sarma, R. J.; Otto, S.; Nitschke, J. R. *Chem. Eur. J.* **2007**, *13*, 9542-9546.

Schultz, D.; Nitschke, J. R. *Angew. Chem. Int. Ed.* **2006**, *45*, 2453-2456.

Searle, M. S.; Griffiths-Jones, S. R.; Skinner-Smith, H. *J. Am. Chem. Soc.* **1999**, *121*, 11615-11620.

Seet, B. T.; Dikic, I.; Zhou, M. M.; Pawson, T. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 473-483.

Severin, K. *Chem. Eur. J.* **2004**, *10*, 2565-2580.

Shinkai, S.; Manabe, O. *Top. Curr. Chem.* **1984**, *121*, 67-104.

Shute, R. E.; Kawai, M.; Rich, D. H. *Tetrahedron* **1988**, *44*, 685-95.

Sidhu, S. S.; Fairbrother, W. J.; Deshayes, K. *ChemBiochem*, **2003**, *4*, 14-25.

Simpson, M. G.; Watson, S. P.; Feeder, N.; Davies, J. E.; Sanders, J. K. M. *Org Lett.* **2000**, *2*, 1435-1438.

Siuti, N.; Roth, M. J.; Mizzen, C. A.; Kelleher, N. L. Pesavento, J. J. *J. Proteome Res.* **2006**, *5*, 233-239.

Sjoback, R.; Nygren, J.; Kubista, M. *Spectrochimica Acta A* **1995**, *51*, L7-L21.

Slomczynska, U.; Hopfinger, A. J.; LeBreton, G. C.; Venton, D. L. *Biopolymers* **1996**, *40*, 617.

Sprangers, R.; Groves, M. R.; Sinning, I.; Sattler, M. *J. Mol. Biol.* **2003**, *327*, 507-520.

Staab, H. A.; Kirrstetter, R. G. H. *Liebigs Ann. Chem.* **1979**, 886.

Stanger, H. E.; Gellman, S. H. *J. Am. Chem. Soc.* **1998**, *120*, 4236-4237.

Stauffer, D. A.; Barrans, R. E.; Dougherty, D. A. *J. Org. Chem.* **1990**, *55*, 2762-2767.

Stolarski, R.; Sitek, A.; Stepinski, J.; Jankowska, M.; Oksman, P.; Temeriusz, A.; Darzynkiewicz, E.; Lonnberg, H.; Shugar, D. *Biochim. Biophys. Acta* **1996**, *1293*, 97-105.

Stotz, C. E.; Topp, E. M. *J. Pharm. Sci.* **2004**, *93*, 2881-2894.

Stotz, C. E.; Borchardt, R. T.; Middaugh, C. R.; Siahaan, T. J.; Vander Velde, D.; Topp, E. M. *J. Peptide Res.* **2004**, *63*, 371-382.

Strahl, B. D.; Allis, C. D. *Nature* **2000**, *403*, 41-45.

Suksai, C.; Gomez, S. F.; Chhabra, A.; Liu, J.; Skepper, J. N.; Tuntulani, T.; Otto, S. *Langmuir* **2006**, *22*, 5994-5997.

Swann, P. G.; Casanova, R. A.; Desai, A.; Frauenhoff, M. M.; Urbancic, M.;

Syud, F. A.; Stanger, H. E.; Gellman, S. H. *J. Am. Chem. Soc.* **2001**, *123*, 8667-8677.

Tatko, C. D.; Waters, M. L. *J. Am. Chem. Soc.* **2002**, *124*, 9372-9373.

Tatko, C. D.; Waters, M. L. *Protein Sci.* **2003**, *12*, 2443-2452.

Taverna, S. D.; Li, Haitao, Ruthenburg, A. J.; Allis, C. D.; Patel, D. J. *Nat. Struct. Mol. Biol.* **2007**, *14*, 1025-1040.

Taverna, S. D.; Ueberheide, B. M.; Liu, Y.; Tackett, A. J.; Diaz, R. L.; Shabanowitz, J.; Chait, B. T.; Hunt, D. F.; Allis, C. D. *Proc. Nat. Acad. Sci.* **2007**, *104*, 2086-2091.

ten Cate, A. T.; Dankers, P. Y. W.; Sijbesma, R. P.; Meijer, E. W. *J. Org. Chem.* **2005**, *70*, 5799-5803.

Thomas, C. E; Kelleher, N. L.; Mizzen, C. A. *J. Proteome Res.* **2006**, *5*, 240-247.

Tie, C.; Gallucci, J. C.; Parquette, J. R. *J. Am. Chem. Soc.* **2006**, *128*, 1162-1171.

Ueno, A.; Tomita, Y.; Osa, T. *Tetrahedron Lett.* **1983**, *24*, 5245-5248.

Ulysse, L.; Chmielewski, J. *Bioorg. Med. Chem. Lett.* **1994**, *4*, 2145-2146.

Ulysse, L.; Cubillos, J.; Chmielewski, J. *J. Am. Chem. Soc.* **1995**, *117*, 8466-8467.

Ura, Y.; Beierle, J. M.; Leman, L. J.; Orgel, L. E.; Ghadiri, M. R. *Science* **2009**, *325*, 73-77.

Van Gerven, P. C. M.; Elemans, J. A. A. W.; Gerritsen, J. W.; Speller, S.; Nolte, R. J. M.; Rowan, A. E. *Chem. Commun*. **2005**, 3535-35537.

Villain, M.; Gaertner, H.; Botti, P. *Eur. J. Org. Chem.* **2003**, 3267-3272.

Voshell, S. M.; Gagné, M. R. *J. Am. Chem. Soc.* **2006**, *128*, 12422-12423.

Wang, G.G.; Allis, C.D.; Chi, P. *Trends in Molec. Med.* **2007**, *13*, 363-372.

Wang, H.; Huang, Z. Q.; Xia, L.; Feng, Q.; Erdjument-Bromage, H.; Strahl, B. D.; Briggs, S. D.; Allis, C. D.; Wong, J.; Tempst, P.; Zhang, Y. *Science* **2001**, *293*, 853-857.

Wang, Y.; Killian, J.; Hamasaki, K.; Rando, R. R. *Biochemistry* **1996**, *35*, 12338-12346.

West, K. R.; Ludlow, R. F.; Corbett, P. T.; Besenius, P.; Mansfeld, F. M.; Cormack, P. A. G.; Sherrington, D. C.; Goodman, J. M.; Stuart, M. C. A.; Otto, S. *J. Am. Chem. Soc.* **2008**, *130*, 10834-10835.

Whiting, A. L.; Neufeld, N. M.; Hof, F. *Tetrahedron Lett.* **2009**, *50*, 7035-7037.

Williams, B. A. R.; Lin, L. Lindsay, S. M.; Chaput, J. C. *J. Am. Chem. Soc.* **2009**, *106*, 6330-6331.

Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1992**. *31*, 1647-1651.

Woll, M. G.; Gellman, S. H. *J. Am. Chem. Soc.* **2004**, *126*, 11172-11174.

Wüthrich, K. *NMR of Proteins and Nucleic Acids*. John Wiley and Sons, Inc.: New York, 1986.

Yana, D.; Shimizu, T.; Hamasaki, K.; Mihara, H.; Ueno, A. *Macromol. Rapid Commun.* **2002**, *23*, 11-15.

Zhang, K.; Siino, J. S.; Jones, P. R.; Yau, P. M.; Bradbury, E. M. *Proteomics* **2004**, *4*, 3765-3774.

Zhang, Y.; Reinberg, D. *Genes & Development* **2001**, *15*, 2343-2360.

Zhang, Z.; Lonnberg, H.; Mikkola, S. *Org. Biomol. Chem.* **2003**, *1*, 3404-3409.

Zhao, T.; Heyduk, T.; Allis, C. D.; Eissenberg, J. C. *J. Biol. Chem.* **2000**, *275*, 28332-28338.

Zubarev, R. A. *Curr. Opin. Biotechnol.* **2004**, *15*, 12-16.